

DOCUMENT RESUME

ED 216 992

SO 014 114

AUTHOR Fels, Rendigs, Ed.; Siegfried, John J., Ed.
TITLE Research on Teaching College Economics: Selected Readings.
INSTITUTION Joint Council on Economic Education, New York, N.Y.
REPORT NO JCEE-312
PUB DATE 82
NOTE 360p.
AVAILABLE FROM Joint Council on Economic Education, 1212 Avenue of the Americas, New York, NY 10036 (\$10.00).
EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Economics Education; *Educational Research; Higher Education; Literature Reviews; Research Methodology; Research Reports; State of the Art Reviews

ABSTRACT

The journal articles reprinted in this publication were selected to aid scholars interested in doing research on the teaching of college economics or in learning about the results of such research. The preface points out that the readings are not intended to provide advice on how to teach economics better. With the exception of one, all of the articles have been previously published in scholarly journals and subjected to their review procedures. The 28 articles are organized into four main sections: surveys of the literature, critiques of research methodology, foundations of research, and research reports. Some examples of the types of articles included are: Male-Female Differences in Economic Education: A Survey; Student Performance and Changes in Learning Technology in Required Courses; The Efficiency of Programmed Learning in Teaching Economics; The Lasting Effects of Introductory Economics Courses; Teacher Effectiveness and Student Performance; Student-to-Student Tutoring in Economics; and Test Information. (RM)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED216992

RESEARCH ON TEACHING COLLEGE ECONOMICS

Selected Readings

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as
received from the person or organization
originating it.
Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL IN MICROFICHE ONLY
HAS BEEN GRANTED BY

Lawrence A. Mayer

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Edited by
Rendigs Fels and
John J. Siegfried

Joint Council on Economic Education

501014 114

RENDIGS FELS is a professor of economics at Vanderbilt University and is one of the founding figures in the study of economic education. He has been treasurer of the American Economic Association since 1976. He was secretary-treasurer from 1970 to 1975, and is a past president of the Southern Economic Association. Professor Fels received his Ph.D. from Harvard University.

JOHN J. SIEGFRIED is a professor of economics and chairman of the department at Vanderbilt University. He was an economist at the Federal Trade Commission in 1975-76 and a senior staff economist for the President's Council of Economic Advisers in 1976-77 while on leave from Vanderbilt. Professor Siegfried received his Ph.D. from the University of Wisconsin.

Contents

PREFACE	vii
I. SURVEYS OF THE LITERATURE	
1. JOHN J. SIEGFRIED AND RENDIGS FELS RESEARCH ON TEACHING COLLEGE ECONOMICS: A SURVEY (<i>Journal of Economic Literature</i> , 1979)	3
2. JOHN J. SIEGFRIED MALE-FEMALE DIFFERENCES IN ECONOMIC EDUCATION: A SURVEY (<i>Journal of Economic Education</i> , 1979)	50
II. CRITIQUES OF RESEARCH METHODOLOGY	
3. BURTON A. WEISBROD RESEARCH ON ECONOMIC EDUCATION: IS IT ASKING THE RIGHT QUESTIONS? (<i>American Economic Review</i> , 1979)	63
4. RICHARD B. MCKENZIE WHERE IS THE ECONOMICS IN ECONOMIC EDUCATION? (<i>Journal of Economic Education</i> , 1977)	71
5. ROBERT J. STAAF STUDENT PERFORMANCE AND CHANGES IN LEARNING TECHNOLOGY IN REQUIRED COURSES (<i>Journal of Economic Education</i> , 1972)	80
6. JUDITH YATES RESEARCH IN ECONOMIC EDUCATION: ARE OUR HORIZONS TOO NARROW? (<i>Journal of Economic Education</i> , 1978)	86
7. W. LEE HANSEN, ALLEN C. KELLEY, AND BURTON A. WEISBROD ECONOMIC EFFICIENCY AND THE DISTRIBUTION OF BENEFITS FROM COLLEGE INSTRUCTION (<i>American Economic Review</i> , 1970)	92
III. FOUNDATIONS OF RESEARCH	
8. ERIC A. HANUSHEK CONCEPTUAL AND EMPIRICAL ISSUES IN THE ESTIMATION OF EDUCATIONAL PRODUCTION FUNCTIONS (<i>Journal of Human Resources</i> , 1979)	101
9. RENDIGS FELS MULTIPLE CHOICE QUESTIONS IN ELEMENTARY EDUCATION (<i>Recent Research in Economics Education</i> , 1970)	139

10.	DARRELL R. LEWIS AND TOR DAHL THE TEST OF UNDERSTANDING IN COLLEGE ECONOMICS AND ITS CONSTRUCT VALIDITY (<i>Journal of Economic Education</i> , 1971)	156
IV. RESEARCH REPORTS		
11.	ELISABETH ALLISON EDUCATIONAL PRODUCTION FUNCTION FOR AN INTRODUCTORY ECONOMICS COURSE (1979; not previously published)	171
12.	RICHARD E. ATTIYEH, G. L. BACH, AND KEITH G. LUMSDEN THE EFFICIENCY OF PROGRAMMED LEARNING IN TEACHING ECONOMICS: THE RESULTS OF A NATIONWIDE EXPERIMENT (<i>American Economic Review</i> , 1969)	195
13.	WILLIAM E. BECKER AND MICHAEL K. SALEMI THE LEARNING AND COST EFFECTIVENESS OF AVT SUPPLEMENTED INSTRUCTION: SPECIFICATION OF LEARNING MODELS (<i>Journal of Economic Education</i> , 1977)	202
14.	JOHN F. CHIZMAR, L. DEAN HIEBERT, AND BERNARD J. MCCARNEY ASSESSING THE IMPACT OF AN INSTRUCTIONAL INNOVATION ON ACHIEVEMENT DIFFERENTIALS; THE CASE OF COMPUTER-ASSISTED INSTRUCTION (<i>Journal of Economic Education</i> , 1977)	218
15.	DAVID G. HARTMAN WHAT DO ECONOMICS MAJORS LEARN? (<i>American Economic Review</i> , 1978)	223
16.	ALLEN C. KELLEY TIPS* AND TECHNICAL CHANGE IN CLASSROOM INSTRUCTION (<i>American Economic Review</i> , 1972) *Teaching Information and Processing System	229
17.	DARRELL R. LEWIS AND CHARLES C. ORVIS A TRAINING SYSTEM FOR GRADUATE STUDENT INSTRUCTORS OF INTRODUCTORY ECONOMICS AT THE UNIVERSITY OF MINNESOTA (<i>Journal of Economic Education</i> , 1973)	236
18.	G. L. BACH AND PHILLIP SAUNDERS ECONOMIC EDUCATION; ASPIRATIONS AND ACHIEVEMENTS (<i>American Economic Review</i> , 1965)	245
19.	PHILLIP SAUNDERS THE LASTING EFFECTS OF INTRODUCTORY ECONOMICS COURSES (<i>Journal of Economic Education</i> , 1980)	273

20. JOHN J. SIEGFRIED
IS TEACHING THE BEST WAY TO LEARN? AN EVALUATION
OF BENEFITS AND COSTS TO UNDERGRADUATE STUDENT
PROCTORS IN ELEMENTARY ECONOMICS (*Southern
Economic Journal*, 1977) 287
21. HOWARD P. TUCKMAN
TEACHER EFFECTIVENESS AND STUDENT PERFORMANCE
(*Journal of Economic Education*, 1975) 294
22. KIM SOSIN AND CAMPBELL R. McCONNELL
THE IMPACT OF INTRODUCTORY ECONOMICS ON
STUDENT PERCEPTIONS OF INCOME DISTRIBUTION (*Journal
of Economic Education*, 1979) 300
23. RALPH D. ELLIOTT, M. EDWIN IRELAND, AND TERESA S. CANNON
AN ANALYSIS OF THE MARGINAL PRODUCTS OF THE ONE-
AND TWO-SEMESTER INTRODUCTORY ECONOMICS COURSE
(*Journal of Economic Education*, 1978) 307
24. ALLEN C. KELLEY AND CAROLINE SWARTZ
STUDENT-TO-STUDENT TUTORING IN ECONOMICS (*Journal
of Economic Education*, 1976) 315
25. MARIAN R. MEINKOTH
TEXTBOOKS AND THE TEACHING OF ECONOMIC
PRINCIPLES (*Journal of Economic Education*, 1971) 319
26. ALLEN C. KELLEY
TEACHING PRINCIPLES OF ECONOMICS: THE JOINT
COUNCIL EXPERIMENTAL ECONOMICS COURSE PROJECT
(*American Economic Review*, 1977) 323
27. WILLIAM F. BARNES
TEST INFORMATION: AN APPLICATION OF THE
ECONOMICS OF SEARCH (*Journal of Economic Education*, 1975) 328
28. SOLOMON W. POLACHEK, THOMAS J. KNEISNER, AND
HENRICK J. HARWOOD
EDUCATIONAL PRODUCTION FUNCTIONS (*Journal of
Educational Statistics*, 1978) 334

Because the articles in this volume have, with one exception, been previously published in scholarly journals and subjected to their review procedures, these readings have been exempted from review by the Publications Committee of the Joint Council on Economic Education. The exception is the article by Elisabeth Allison which, although not previously published, is regarded by researchers in economic education as a significant contribution to their field.

Preface

JOHN J. SIEGFRIED AND RENDIGS FELS

I

We have selected these readings to aid scholars interested in doing research on the teaching of college economics or in learning about the results of such research, rather than to provide advice on how to teach economics better. While the ultimate purpose of all this research is to improve the teaching and learning process, its immediate purpose is to accumulate more knowledge about what happens in the classroom. Most of the papers in this book are therefore mainly methodological or not directly applicable to teaching. A few do contain information relevant to classroom procedures. However, much richer sources of specific advice about how to teach economics are *Resource Manual for Teacher Training Programs in Economics* (1978)*, edited by Phillip Saunders, Arthur L. Welsh, and W. Lee Hansen, as well as other materials used in the Teacher Training Program of the Joint Council on Economic Education.

In one sense, research on economic education is an old field, in another sense it is quite new. When Laurence E. Leamer wrote his "Brief History of Economics in General Education" for the *American Economic Review* (1950), he found a voluminous literature already in existence. But previous authors had written so casually as to suggest that no one had ever considered the subject before and that no one would bother to go on from where they had left off. Leamer ended with a plea for economists to "build a cumulative literature on their subject" and added that "progress in economic education is hardly to be made in the next sixty years if we are to persist, as we have in the past, in retracing previously trod paths" (p. 33).

A decade went by before Leamer's plea began to be heeded. Simon N. Whitney pioneered with his "Measuring the Success of the Elementary Course" (1960), but his was

an isolated effort. In 1961, the American Economic Association (AEA) began to publish material from its sessions devoted to economic education (see the annual *Papers and Proceedings* issues of the *American Economic Review*.) In the early years these papers mainly consisted of informal observations rather than findings based on genuine research. In the mid-1960s, however, several developments helped stimulate the beginnings of a cumulative literature: the *Test of Economic Understanding* (TEU) was published; a television series on "The American Economy" was shown nationwide; programmed instruction (PI) appeared; the *Test of Understanding in College Economics* (TUCE) was developed. Their impact was enhanced by the coming of age of the computer, by the decision of the Joint Council on Economic Education to add a college-and-university program to its existing activities in lower schools, and by the thrust to research given by the AEA's Economic Education Committee under the leadership of G. L. Bach.

Although the TEU was designed for high schools, it was nevertheless used to conduct research on the effectiveness of the TV presentations of "The American Economy," for which a number of institutions offered college credit to participants (Saunders, 1964; McConnell and Felton, 1964; Bach and Saunders, 1965, reading 18 below). The paper by Bach and Saunders was the first major research report on economic education ever to be published, at least in this country, and it is still an interesting document.

Although programmed instruction is no longer so greatly in vogue, its advent in the 1960s stimulated much enthusiasm. A number of books on PI appeared, and research was undertaken on its effectiveness. A

*All publications mentioned here are cited in full in the References at the end of this Preface or are reprinted in this book.

notable result was the second major research paper on economic education. It was written by Attiyeh, Bach, and Lumsden (1969), and is reprinted below. The authors were able to use a preliminary version of the TUCE, a test suitable for use as a measuring instrument in colleges. TUCE sparked a considerable expansion of research activity. The major paper in this wave was Saunders' study of the lasting effects of economic education, also reprinted here.

In the late 1960s, Keith G. Lumsden edited two books (1967; 1970) of papers presented at two conferences he organized. They were the first books entirely devoted to research on the teaching of college economics.

Another big step was taken in 1969, when the semi-annual *Journal of Economic Education* (JEE) began publication. In publishing the JEE, the initial emphasis of the Joint Council on Economic Education (JCEE) was on encouraging research rather than providing an outlet for worthy studies that might not otherwise get into print. Ultimately the JEE did elicit an enlarged body of worthwhile research, though progress was slow at first.

In 1973, we thought the time might be ripe for a book of reprints of the best work done so far. We drew up a proposed table of contents and showed it to friends, but their reaction discouraged us from proceeding. At about the same time, the editor of the *Journal of Economic Literature* (JEL) commissioned us to write an article surveying the literature devoted to research on teaching college economics. Fortunately, both the book of readings and the survey were delayed, for in the next five years the sophistication of the research increased markedly. As we wrote the survey—the first article reprinted below—we became convinced that the time had come to publish the volume of readings we had proposed several years earlier. The dates of the material reprinted in this book testify to what had happened to the quality of research in economic education in the interim: of the 28 papers included here, 19 were published after mid-1975.

II

Neither in our survey article nor in this book do we cover the literature on precollege economic education. There are several

reasons. Research on economic education is carried out primarily by economists in colleges and universities; consequently it tends to be about education at such institutions. Our own knowledge is primarily about those levels. The readers in whom we can expect to arouse the most interest are college teachers. And in any event, keeping up with the college and university literature these days is a job in itself. Those interested in the precollege literature should consult the survey by Dawson (1976). Earlier surveys were conducted by Leamer (1950), Fels (1969), Lewis and Orvis (1971), and Wells (1974).

Our survey article in reality constitutes the introduction to this book. It summarizes and discusses most of the papers selected for inclusion and obviates the need for us to say anything more about them. We have, however, included a few items we did not, for one reason or another, mention in the survey. William F. Barnes's "Test Information: An Application of the Economics of Search," as do most of the other papers, provides answers to the question raised by Burton F. Weisbrod, viz, Why should economists do the research on economics education rather than leaving the subject to psychologists and educationists? Robert J. Staaf's "Student Performance and Changes in Learning Technology in Required Courses" gives some of the results reported in a book we discuss in our survey. "Student-to-Student Tutoring in Economics," by Allen C. Kelley and Caroline Swartz, and "Textbooks and the Teaching of Economic Principles," by Marion Meinkoth, are relevant to our subject although they could not conveniently be fitted into the survey.

There are four sections below. The first, "Survey of the Literature," consists of two papers: our survey for the JEL and a paper by one of us, Siegfried, originally written as part of our survey but omitted because of space constraints. Section II, "Critiques of Research Methodology," contains five commentaries on research in economics education. Section III, "Research Foundations," includes a paper on educational production functions by Eric A. Hanushek. Although it is not directly concerned with the teaching of economics in college, it is valuable for anyone concerned with research on the topic. Section III also includes two papers on multiple choice tests in general and the TUCE in particular. Section IV, "Research

Reports," makes up about half the book. These reports present substantive findings on which we believe and hope that future investigators can build their work.

III

We are deeply indebted to the Joint Council on Economic Education not just for publishing and subsidizing this book but, more importantly, encouraging and supporting research on economic education by publishing the *Journal of Economic Education*, sponsoring sessions at meetings of economics associations, sponsoring the *Test of Understanding in College Economics* as well as other tests, stimulating investigations and publishing the results in *Research Papers in Economic Education* (Welsh, ed. 1972), and through countless other activities. Everyone concerned with economic education is also indebted to all those whose financial support has made the work of the Joint Council possible.

REFERENCES

Note: This list omits papers mentioned in the Foreword that are reprinted in this book.

Dawson, George. "Research in Economics Education at the Pre-College Level." Mimeo, 1976.
Fels, Rendigs. "Hard Research on a Soft Subject:

Hypothesis-Testing in Economic Education." *Southern Econ. J.*, July 1969.

Leamer, Lurence E. "A Brief History of Economics in General Education." *Amer. Econ. Rev.*, December 1950, Part 2, Supplement.

Lewis, Darrell R., and Charles C. Orvis. *Research in Economic Education: A Review, Bibliography and Abstracts*. New York: Joint Council on Economic Education, 1971.

Lumsden, Keith G., ed. *New Developments in the Teaching of Economics*. Englewood Cliffs, N.J.: Prentice-Hall, 1967.

—. *Recent Research in Economics Education*. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

McConnell, Campbell R., and John R. Felton. "A Controlled Evaluation of The American Economy." *Amer. Econ. Rev.*, June 1964.

Saunders, Phillip. "The Effectiveness of The American Economy' in Training Secondary School Teachers." *Amer. Econ. Rev.*, June 1964.

Saunders, Phillip; Arthur L. Welsh; and W. Lee, Hansen. *Resource Manual for Teacher Training Programs*. New York: Joint Council on Economic Education, 1978.

Wells, Stuart. "The Effectiveness of Economics Education: A Survey of Research Findings." Faculty of Commerce, University of British Columbia, Working Paper Number 331, November 1974.

Welsh, Arthur L., ed. *Research Papers in Economic Education*. New York: Joint Council on Economic Education, 1972.

Whitney, Simon N. "Measuring the Success of the Elementary Course." *Amer. Econ. Rev.*, March 1960.

I
Surveys of the Literature

Research on Teaching College Economics: A Survey

By JOHN J. SIEGFRIED AND RENDIGS FELS

Vanderbilt University

We are indebted to Elisabeth Allison, G. L. Bach, William Becker, Frank Bonello, Kenneth Boulding, Stephen Buckles, J. R. Clark, George Dawson, Daniel Fusfeld, Malcolm Getz, W. Lee Hansen, Robert Heilbroner, Robert Highsmith, Cliff Huang, Thomas Johnson, Allen Kelley, Darrell Lewis, Michael MacDowell, Campbell McConnell, Richard McKenzie, David Morawetz, Donald Paden, Phillip Saunders, Alex Scott, Howard Tuckman, John Vahaly, Henry Villard, Burton Weisbrod, Arthur Welsh, and Thomas Zak for comments on an earlier draft; to Thomas Overstreet and James Lewek for research assistance; to the members of Economics 380, Kaye James, Noel Lim, Katherine Maddox, Hal McClure, Mary Ann Meiners, and George Nomikos, for papers and discussions on economics education; to Violet Sikes for typing; and to Marjorie Churchill for editorial assistance.

I. Introduction

DURING the past fifteen years, an extensive research literature has developed on the subject commonly called economic education (not to be confused with the economics of education).¹ Economists with established reputations have made

¹ The economics of education is concerned primarily with the benefits, costs, production, and financing of the dissemination of knowledge of any and all subjects. Economic education has similar concerns regarding the dissemination of knowledge about the single subject of economics. Most of the research evaluates alternative strategies for teaching economics successfully. The field is sometimes called economics education, a term that is better English usage.

contributions, the *Journal of Economic Education* has completed almost a decade of publication, and some economists regard the field as their primary specialty.²

This survey is as diverse as the research it reports, but several themes recur. (1) In preparing this paper, we became impressed by the usefulness of the tools of economic theory and econometrics for analyzing teaching problems. Burton F. Weisbrod has raised the question: Why should economists investigate teaching problems at all? Why not leave the subject

² In the 1978 *Directory* of the American Economic Association, 45 economists indicated that economic education was their primary specialty [4, 1978, pp. 426-31].

to educational specialists and psychologists?³ [171, 1979]. The most telling answer lies in the nature of the work already done. The economist brings to research on teaching certain skills and ways of thinking indispensable for sophisticated investigation of the subject. (2) Another theme is diversity. Different students learn in different ways. A menu of techniques may dominate any single one. (3) From the contributions of psychologists comes the theme that instruction targeted at specific objectives may be better than instruction not so targeted. (4) We shall be concerned with the extent to which training and experience lead to better teaching, particularly whether training in teaching as distinct from training in subject matter has anything to contribute. (5) We shall also be concerned with the search for innovative teaching methods that work.

A. Omissions

The present survey concentrates on college teaching, particularly the elementary course (because most of the research has been on it). We pay only slight attention to high school economics and none to pre-high school teaching, graduate instruction (except for training of college economics teachers), and informal economics education.⁴ With minor exceptions, we confine ourselves to literature published in

economics journals and books by economists. Most of the work is in the *Journal of Economic Education* and the annual *Papers and Proceedings* of the American Economic Association. We focus on research findings in contrast to descriptions of courses, teaching methods, and curricula.

We have omitted certain topics partly due to space constraints and partly because they do not fit well with the main focus of the survey. We regret the omission of radical economics. Radical economists have criticized orthodox economics teaching severely and have published discussions of alternative approaches.⁵ But to the best of our knowledge, they have not done enough research on the results to warrant inclusion here. For similar reasons, we have omitted discussion of the experimental elementary courses sponsored by the Joint Council on Economic Education (JCEE).⁶ Also omitted is a discussion of sex differences in economics education, which would report that females enter college at a disadvantage (apparently for cultural reasons), which tends to persist but not to worsen during introductory economics.⁷

Three decades ago a subcommittee of the American Economic Association reported that "the content of the elementary course has expanded beyond all possibility of adequate comprehension and assimilation by a student in one year of three class hours a week" [162, Horace Taylor, 1950, p. 56]. This kind of criticism

³ A prior question is whether fruitful research on such problems is possible by anybody. Casual empiricism by the older of us suggests that 20 years ago large numbers of economists and faculty members in liberal arts colleges generally would have answered with a resounding no. It is likely that many of them still feel that way. The proper answer is that any important activity like college teaching is appropriate for research. The only questions are when a subject becomes ripe for fruitful investigation and whether the likely results justify the cost. Thirty or forty years ago an answer of "not yet" was plausible, but no longer.

⁴ Those interested in precollege economic education are referred to Donald R. Wentworth, W. Lee Hansen, and Sharryl H. Hawke [175, 1977] and W. Lee Hansen *et al.* [59, 1977].

⁵ The first draft of this survey discussed Richard C. Edwards and Arthur MacEwan [41, 1970]; Michael Zweig [179, 1972]; John G. Gurley [56, 1975]; Michael Meeropol [105, 1975]; and Robert Buchele and William Lazonick [26, 1975]. See also Martin Bronfenbrenner [24, 1970, p. 765].

⁶ See Kenneth and Elise Boulding [22, 1974]; Rendigs Fels [49, 1974]; Richard H. Leftwich and Ansel M. Sharp [89, 1974]; Phillip Saunders [133, 1975]; and Barbara and Howard Tuckman [165, 1975].

⁷ For a detailed survey of the literature evaluating the relationship between sex and learning economics see Siegfried [142, 1979], which was a portion of the first draft of this article.

has been repeated again and again.⁸ According to George Stigler, "The watered-down encyclopedia which constitutes the present course in beginning college economics does not teach the student how to think on economic questions. The brief exposure to each of a vast array of techniques and problems leaves the student with no basic economic logic with which to analyze the economic questions he will face as a citizen. The student will memorize a few facts, diagrams, and policy recommendations, and ten years later will be as untutored in economics as the day he entered the class" [159, 1963, p. 657]. Such criticism is based on an assumption about what the typical elementary course is like. The assumption may be true, but has never been verified. We shall say no more about it.

Whereas the introductory course has been the subject of much discussion, criticism, and research, the undergraduate major has drawn little attention. An exception is David Hartman's study [64, 1978; 65, 1978].⁹ According to Hartman, the goal implicit in the general examinations given to undergraduate majors at Harvard is skill in micro theory, macro theory, and the use of each "to answer real world questions" [64, 1978, p. 18] (see also [65, 1978, p. 87]). He concluded that the introductory course is the high point of the undergraduate major at Harvard and that the conventional course in inter-

mediate micro theory "does not appear to have much impact on either students' knowledge of micro theory or their ability to apply it" [65, 1978, p. 90]. He found some evidence that a policy-oriented macro course has an impact on what students know about macro theory, but it is not strong enough to pass a significance test at conventional levels of acceptance.

B. The Production Function for Economics Education

In the rest of the survey, we organize the massive literature on teaching methods and techniques around the typical components in production function analysis. Production function studies can help determine whether or not economics instruction is efficient. This has important policy implications, especially for times of austere instructional budgets, since inefficiency means that it is possible to increase school outputs without additional inputs.¹⁰ The vast majority of the research on economics education has been concerned with evaluating teaching methods. The implied theoretical model is a production function that shifts as a consequence of a particular technique.¹¹ The conceptualization of these empirical studies, however, has been disorganized.

The absence of an established theory of learning has made the specification of

⁸ For other examples of the standard criticism, see Fels [46, 1955; 49, 1974], George L. Bach [11, 1967; 12, 1976], Leftwich [88, 1976], Gurley [56, 1975], Hansen [57, 1975], Leftwich and Sharp [89, 1974], Saunders [133, 1975], and Barbara and Howard Tuckman [165, 1975].

⁹ Other recent exceptions, albeit with little research content, are David Morawetz [107, 1976; 109, 1978] and Bronfenbrenner [25, 1978]. According to Morawetz, "The aim of an economics education is to train students to become economists" [107, 1976, p. 1]. Such a goal seems irrelevant to the vast majority of undergraduates in American colleges and universities. Bronfenbrenner says, "I prefer economics to become more frequently a 'junior Ph.D. major'" [25, 1978, p. 22].

¹⁰ There is, however, a fundamental statistical and interpretive issue: Most production function studies are based on observed behavior, in which case the estimated relationships will not map the production frontier unless all producers are generating maximum output for given inputs. Ironically, we are usually engaged in such analyses because we suspect that production is not occurring on the frontier.

¹¹ Most of the difficulties in modeling and estimating economics education production functions are common to the specification and estimation of general educational production functions. However, in the latter case more attention has been devoted to the impact of sociodemographic characteristics than to teaching technologies on school performance. Comprehensive surveys of this literature are included in Elchanan Cohn [32, 1979] and Richard Murnane [112, 1975].

production functions difficult. In general, linear additive models have been specified. There has been little concern with such issues as simultaneity (if student interest is an output, does it not feed back into cognitive understanding?), functional form (are there interactions among the independent variables?), and the statistical techniques employed (many dependent variables are limited, even dichotomous, making ordinary least squares regression analysis inappropriate). The studies have usually arisen from an instructor's initiative to try a new teaching technique and typically involve a comparison of the course before and after or with and without the new innovation. Only a limited number of the studies have used broad-based samples, and very few have considered whether the benefits are worth the costs.

To evaluate innovative instructional techniques it is necessary to hold other things, especially the level of inputs, constant. William I. Davisson and Frank J. Bonello [40, 1976] describe a useful taxonomy for organizing research on the production function for learning college economics. Their approach is superior to the *ad hoc* theorizing that characterizes most of the economics education literature. Davisson and Bonello identify three separate categories of inputs: human capital (*SAT* scores, grade point average, and pretest scores); utilization rates (time spent on the course by students); and technology—the efficiency with which effort is transformed into cognitive achievement (lecturer effectiveness, text effectiveness, etc.). This distinction draws attention to the potentially defective conclusions that may arise from simple comparisons of test scores between control and experimental groups, or even a comparison of control and experimental groups' performance that controls for human capital characteristics. A shift in the production function as a consequence of some alternative tech-

nology can be detected correctly only if input utilization rates are held constant. Otherwise, performance comparisons consist of output at different levels of inputs on potentially different production functions, and it is impossible to disaggregate the effect of changes in the level of inputs from changes in the rate at which inputs are transformed into outputs.

The production function involves multiple outputs as well as multiple inputs. According to Judith Yates, "the objectives of the education system must always be borne in mind [in research on economics education]. These objectives are often ill-defined and may vary significantly among institutions. . . . At the risk of overlooking important ones, a number of objectives can be listed: students' growth and development with regard to skills (which can include application, critical thinking, creativity, motor skills, etc., as well as comprehension and understanding); their social development (including leadership, communication, interpersonal relations, etc.); their acquisition of vocational skills, and so on." Furthermore, "The contribution of a particular technique or teaching method to the process of learning how to learn rather than to the specific output of precisely what was learned may be its most important attribute" [178, 1978, p. 13]. Other outputs are increased sophistication with respect to values and attitudes and changes in ideological orientation. Colleges also typically serve a screening role for graduate schools and the labor market.

II. Measuring Outputs

Those doing research on college teaching of economics are only beginning to come to grips with the multiplicity of objectives. In practice, they have measured output in three principal ways: examinations (usually objective tests), student evaluation questionnaires, and a kind of market test using number of majors attracted

or enrollment. The unsolved problem of measuring all the outputs and assigning weights to them has been pointed out by Yates and others as limitations on the research done. A related problem, which has been receiving attention in recent years, arises from the fact that different teaching methods have different impacts on different students, so that the distribution of outputs needs to be measured. This section is organized around five dimensions of output: cognitive performance, student attitudes toward the subject and the educational process, the impact of understanding and attitudes on subsequent behavior, changes in values, and the distribution of benefits.

A. Measurement of Learning

The bulk of research on college teaching of economics is based on the implicit assumption that "the major criterion or objective of economic education is a very narrowly defined concept of learning, related in some way to student performance as measured by achievement or course grade" [178, Yates, 1978, p. 12]. Availability of nationally normed, validated objective tests from which the data can be analyzed with standard statistical tools has encouraged this procedure.

A number of objective tests are available for research on economics education. (1) The *Test of Economic Understanding (TEU)* is a 50-item multiple choice test of what every high school graduate should know about economics [71, JCEE, 1963]. It has recently been revised [154, John C. Soper, 1979]. The new version is called the *Test of Economic Literacy (TEL)*.¹³ For a time the *TEU* was used to measure learning in the college elementary course, a purpose for which it is unduly simple. (2) The elementary economics test in the Educational Testing Service's College

Level Examination Program is a 90-minute 100-item multiple choice test designed to enable colleges to determine whether transfer students and students who have learned economics on their own are entitled to credit for the elementary course and/or admission to advanced undergraduate courses. It concentrates heavily on theory. Separate micro and macro versions are available. Because of high cost and tight security regulations, it has been used for only one research project that we know of, namely John J. Siegfried [141, 1977]. (3) By far the most widely used instrument is the *Test of Understanding in College Economics (TUCE)* [72, JCEE, 1968], which has been used in at least 71 studies.¹³ For other discussions of the *TUCE*, see Fels [48, 1970], Arthur L. Welsh and Fels [174, 1969], Darrell R. Lewis and Tor Dahl [91, 1971], and T. R. Swartz, F. J. Bonello, and W. I. Davisson [161, 1977]. (4) The *Test of Economic Comprehension (TEC)*, which is similar to the *TUCE* in content and construction, has been used mainly in Great Britain [9, Richard Attiyeh and Keith Lumsden, 1971].

The *TUCE* is in two parts, each of which has two alternative forms of 33 multiple choice questions. Part I is intended to cover the content of the typical first semester of the college course, *i.e.*, macroeconomics plus some basic micro concepts. Part II is intended for the typical second semester: the theory of the firm, marginal productivity analysis, international economics, and comparative systems.¹⁴ Since the two forms for each part are approximately equal in difficulty, one

¹³ For this information, we are indebted to John Vahaly.

¹⁴ Supply and demand analysis is included in both Part I (bare elements) and Part II. In addition to the four forms described above, a "hybrid" test consisting of 33 questions selected from both Part I and Part II was devised for Saunders's lasting-effects study discussed below. See Saunders and Welsh [136, 1975] and Saunders [129, 1971; 132, 1973].

¹³ The *TEL* is meant for grades 11 and 12. The Joint Council on Economic Education, its publisher, also has tests for grades 2-3, 4-6, and 7-9.

can be given at the beginning of the semester, the other at the end, to measure the amount of value added. Although in subject matter content the specifications were meant to conform to the typical college course, they depart from it with respect to kinds of questions, having approximately one-third each in the categories of recognition-and-understanding, simple applications, and complex applications. (See Fels [47, 1967] and Welsh and Fels [174, 1969]). The committee responsible for the test hoped to influence economics instructors to put more emphasis on applications and less on memorization and abstract theory. National norms were established using data from 50 colleges.

The heavy dependence of economics education research on the *TUCE* calls for appraisal of its strengths and weaknesses. On the plus side, the quality of the test is considerably higher than the tests that can be constructed by individual research workers.¹⁵ Constructing a reasonably satisfactory test is enormously difficult and expensive. The existence of matched pairs of tests, making possible measurement of value added in controlled experiments, and the availability of norming data for comparison purposes are legitimate reasons for the popularity of the *TUCE*. It was tested for validity [91, Lewis and Dahl, 1971] and found to be an effective discriminator of students with high and low levels of ability and a good measure of prior ability and analytical skills [28, Stephen G. Buckles and Welsh, 1972]. Lewis and Dahl found that its simple application questions were correlated with

critical thinking skills as measured by the *Watson Glaser Critical Thinking Appraisal Test*.

The shortcomings of the *TUCE*, however, are considerable. In the first place, no general purpose test is likely to conform exactly to the purposes and content of any particular course. In the second place, not all the questions are satisfactory. (The *TUCE* is currently undergoing revision to replace unsatisfactory and outdated questions.) In the third place, no multiple choice test can measure all the objectives of elementary instruction.¹⁶

The correlation between scores on the best possible multiple choice test and grades on essay examinations varies according to the subject—high for mathematics, low for English literature, with economics an intermediate case. We do not know of any study directed at the specific question of how *TUCE* test scores correlate with essay tests built to the same specifications. H. Tuckman found a negative simple correlation between a ten-question version of the macro *TUCE* and course grades [166, 1975]. P. F. Labinski at a community college found an r^2 of .22 between the *TUCE* and course grades. He argued that the objectives of economics instruction at two-year colleges differ from four-year colleges [86, 1978]. Inasmuch as the *TUCE* was designed to measure not so much the objectives of four-year colleges themselves but what a committee thought those objectives should be, it is not surprising that two-year colleges with still other objectives would find the *TUCE* a poor measure of their output. Elisabeth Allison, in a Harvard Working Paper that estimates production functions for economics education, reports correlations ranging over a three-year period from .74 to .85 between scores on a 60-minute, 40-question multiple-choice test

¹⁵ Despite the widespread use of multiple choice questions for grading purposes, most economists are not familiar with the elementary principles of test construction, e.g., that specifications for the test must be drawn up in advance, that questions need to be pretested on seven or eight hundred students, that the right response must be shown in the data to be chosen more often by good students than by poor students and vice versa for the distractors, and that questions need to be edited by a psychometrician (among other reasons to avoid tipoffs to testwise students). For details, see Fels [48, 1970].

¹⁶ T. R. Swartz, Bonello, and Davisson have documented the shortcomings of the *TUCE* as a measure of cognitive achievement for Notre Dame [161, 1977].

(parts of which are from *TUCE*) and a two-hour essay portion of the same three-hour final examination at Harvard [3, 1977, p. 15].¹⁷

Future progress in economics education will require a computerized national test bank consisting of thousands of carefully edited multiple choice questions with data proving that they "work." The questions need to be classified not only by subject and type of question but also by kind of course or courses for which they are useful (high school, elementary college, intermediate theory, etc.).

The availability of two matched forms of the *TUCE* has led to pre- and post-course testing, which permits several forms of the output measure to be specified: (1) absolute achievement—the post-test score; (2) absolute improvement—the difference between the post-test and the pre-test score; (3) percentage improvement—absolute improvement divided by the pre-test score; and (4) gap closing measure—absolute improvement divided by the potential gain in score (which is the difference between the perfect score and the pre-test score). The absolute achievement score reflects the level of understanding at a point in time. It is a stock measure. The absolute improvement score measures the increment of learning during a course.¹⁸ An alternative is to use

the post-test achievement score and control for initial economic understanding by including the pre-test score as an independent variable in the regression analysis. The percentage improvement and gap-closing measures were developed because there may not be a constant difficulty of learning throughout the spectrum from total ignorance to total mastery; thus any aggregate measure of success should measure improvements by students on the basis of the difficulty of achieving them [177, Simon Whitney, 1960]. The percentage improvement form implies that it is more difficult for poorer students to improve their scores by a given absolute amount. The gap-closing measure implies the opposite, that it is more difficult to improve a score by a given amount if one starts at a higher level of mastery (because only the most difficult material remains to be mastered) [155, Soper and Richard Thornton, 1976, pp. 86–87].

At first glance absolute improvement, percentage improvement, and gap-closing measures of output may appear subject to ceiling or floor effects. Random guessing would generate about a 25 percent score on a multiple choice test with three distractors in each question. Since the average score in the *TUCE* national norm was 57 percent, floor effects seem more likely than ceiling problems. William Becker and Michael Salemi [18, 1977] used a nonlinear model to deal with ceiling effects. More research is required on this issue because the existence of an upper or lower boundary tends to make the dependent variable inappropriate for ordinary least squares estimation.

stock generally report positive correlations between end-of-course student performance and male sex. Studies of the flow generally do not. Only when the distinction between stock and flow became recognized was the correct conclusion drawn: the evidence shows only that females start the introductory economics course at a disadvantage, not that they are at a disadvantage in learning economics during the course.

¹⁷ The major study in the United Kingdom (The Economics Education Project) carried out between 1969 and 1973 analyzed the correlation between *TEC* scores and grades awarded at both university and high school levels. At high school level the R^2 between final grades as awarded by the Examination Boards varied from .05 to .26 (sample sizes: 1970, 4,254 students; 1973, 3,335 students). At university level the R^2 was significantly greater than zero (ranging from .05 to .49 for 27 out of 38 courses with a total of 2,823 students. The mean R^2 was .12. (We are indebted to Alex Scott for this footnote. These results will be in a forthcoming book by Lumsden, Attiyeh, and Scott [97].

¹⁸ Failure to consider explicitly whether the stock or the flow is to be measured is a common fault. For instance, most studies of economics education include sex as a binary variable but are not explicit about whether it is the stock or the flow that is associated with sex. Studies implicitly concerned with the

Incentives seem to make a difference. Sometimes tests used to measure output in an experiment count toward grades, sometimes not. William Wehrs found that the greater incentives of counting the *TUCE* when used as a post-test made a 12 percent difference in absolute score (and a gain difference of 40 percent) after holding constant the pre-*TUCE* score, high school rank, age, and sex [168, 1978]. Only a quarter of the instructors participating in norming the *TUCE* used the post-test results for grading, biasing the norms downward. But the scores used for the pre-test norms did not count toward grades at all, biasing improvement upward.

Most learning technology improvements have been adopted in introductory courses. Richard McKenzie and Robert Staaf argue that introductory courses, which are frequently part of "distribution requirements," are likely to be inferior goods [104, 1974, pp. 30-33]. The intent of introducing improvements in teaching technology may be to induce a substitution effect toward the discipline, but there is also an income effect associated with the relative price change (in terms of reduced student effort to achieve a fixed learning or grade level) brought about by improved technology in a single discipline. The benefits from improvements in teaching introductory economics may manifest themselves in released time devoted to studying other subjects or consumed in leisure. Allen Kelley found empirical support for this proposition [83, 1975]. If this occurs, the technology change may be deemed successful unless we value the alternative uses of time at zero.

B. *Measurement of Student Attitudes*

Student evaluations of courses and professors have come into widespread use in American colleges and universities over the past decade. Three types of student

attitudes are likely to be of interest: (1) attitudes toward policy issues, (2) attitudes toward economics, and (3) opinions regarding the quality of instruction. The first will be considered in the subsection on values. The second was studied by David Ramsett, Jerry Johnson, and Curtis Adams using data from three small midwestern universities [122, 1973]. They found that those students who were more favorably disposed toward the subject of economics did better on the post-course *TUCE*, holding pre-*TUCE* and sociodemographic factors constant. Lewis Karstensson and Richard Vedder confirmed these findings with a similar analysis of Ohio University students, finding that students whose interest in economics grew the most during the semester also learned the most [74, 1974]. However, there was no attempt to control for study time.

The principal controversies surrounding student evaluations of teaching arise from the various uses to which they can be put. Some have argued that course evaluations are undesirable because they do not adequately represent student cognitive achievement. Miriam Rodin and Burton Rodin, using 12 calculus classes, reported that students who learned the most rated their instructors the worst [124, 1973]. Similar findings were reported by Dennis Capozza, who regressed student ratings of instructors on the increase in *TEU* scores and the average grade of eight different economic classes [29, 1973]. He found that higher course evaluations were associated with higher grades and less gain in *TEU* score. Studies by Soper, using data from the University of Missouri [151, 1973], and by Attiyeh and Lumsden, using a sample of approximately 4,700 British students [9, 1971], yielded similar results.¹⁹

¹⁹ Both the Rodin and Rodin [124, 1973] and the Soper [151, 1973] studies were of large lecture classes with discussion groups led by graduate teaching assistants. The instructor evaluations were of the

On the other side of the issue are data from Saunders [131, 1972], which appear to support the validity of student evaluations as measures of cognitive achievement in three of four tests, and from Peter Sloane [149, 1972]. A more recent study by W. Douglas Morgan and Jon Vasché [109, 1978] also concludes that students can and do recognize good teaching. They estimated the marginal product of various teaching assistants (TA's) in a large introductory-level macroeconomics course at the University of California, Santa Barbara, by regressing final student grade performance on typical control variables and binary variables for specific teaching assistants. They then correlated the student course evaluations of the TA's with their impacts on final course grades, which the TA's did not control. There was a perfect rank correlation of evaluation scores with the estimated marginal products on (1) preparation for class, (2) communication skills, and (3) ability to respond to questions. Ratings on knowledge of the subject were positively but not perfectly correlated with marginal products.

Douglas Needham has reconciled the conflicting empirical relationships between learning, course evaluations, and expected grades [113, 1978]. Using a model of student time allocation in which students equate the marginal grade productivity of time devoted to each activity, Needham shows that the theoretical expectations relating grade levels to course evaluations and relating student learning to course evaluations are ambiguous because student time allocation decisions depend on the *rate* of transformation of ef-

fort into learning and the *rate* of transformation of learning into grades, rather than the absolute levels of learning and grades.

The debate on whether student course evaluations measure cognitive achievement well may overlook the multiple dimensions of teaching. If parties in the instructional process value the various dimensions differently, there may well be reason for an assessment of instructional characteristics aside from cognitive achievement. The resolution of this question then hinges on the assignment of values to different aspects of educational output. Some of the parties to the process may value "entertainment" highly and desire it to be a part of their education. Even if this is what course evaluations assess, they may be warranted.²⁰

Students' assessment of instruction can be obtained either through systematic course evaluations, informal student opinion, or carefully structured intensive interviews. Although systematic course evaluations have the advantage of a larger sample, they may be subject to greater measurement error than intensive interviews because students feel less responsible in responding. But since systematic course evaluations are usually anonymous, they might be less biased.

The interpretation of student evaluations is also important. By far the most controversial issue is whether instructors can "buy" higher evaluations by lowering the (effort) price to students of achieving a given grade. McKenzie [102, 1975] and Paul Kipps [83, 1975] provide theoretical analyses employing price theory. McKenzie shows how an instructor can increase the average course grade by a parallel

teaching assistants. Kelley found that course evaluations were not strongly influenced by the performance of individual teaching assistants [81, 1972]. If this is correct, then lower evaluations of assistants whose students did worse on cognitive achievement may be spurious, the major effect on their performance arising from the lectures and other differences among them, which were not controlled in the Rodin and Rodin and Soper studies.

²⁰ However, Frank Costin, William Greenough, and Robert Menges in their comprehensive assessment of student course evaluations (mostly in teaching psychology) conclude that "existing data . . . do not permit a conclusion that sheer 'entertainment' is what makes students perceive a teacher as a 'good' one" [35, 1973, p. 51].

shift in the leisure-grade transformation curve, a shift favoring the better students, or a shift favoring the poorer students. If course evaluations improve as students do better in the class relative to their expectations, a relaxation of grading standards is consistent with either a positive, neutral, or negative correlation with individual student course evaluations. If the worst students are the ones who are pleasantly surprised by grading standards, they will rate the instructor higher than expected and, *ceteris paribus*, create a negative overall correlation between evaluations and grades for the course. But if instructors "buy" evaluations by raising grades, the positive correlation should appear in cross-course comparisons [76, James Kau and Paul Rubin, 1976].

These expectations seem to be borne out in the empirical studies. Kelley regressed course and professor evaluations on expected course grade, student ability, and other standard control variables for 258 students at the University of Wisconsin [80, 1972]. He found that the coefficient on students' expected course grade was significantly greater than zero, but its magnitude was small. It would have taken an enormous change in the instructor's overall grading standards to generate a trivial movement in his course evaluation scores.

Studies of the association between perceived grading standards and course evaluations across classes have found the expected relationship—easier grading is positively correlated with evaluations. In an analysis of 339 sections of social science courses at Central Michigan University, Alan Nichols and Soper found that students' expected course grade was a significant (both statistically and practically) determinant of course evaluations, holding course environment characteristics constant [115, 1972]. Class size and course level were not related to course evaluations. Rolf Mirus [106, 1973], in a regres-

sion study of 122 courses at the University of Alberta, confirmed the Nichols and Soper findings. He found that expected course grade was significantly related to course evaluations, but class size, type of course, the evaluation response rate (fraction of enrolled students who completed the evaluation), or whether the course was required did not influence evaluations. According to Mirus's findings, "a professor who, compared to his colleagues, makes the class expect a 1 point higher grade can improve his own evaluation by .85 of a point" [106, 1973, p. 36]. Both course evaluation and student grade were measured on a five-point scale. In a study of 201 classes in the College of Business Administration at the University of Georgia, Kau and Rubin reported similar findings [76, 1976]. Class level, required course, and expected grade were all statistically significant determinants of course evaluations. The average class grade point average, class size, and the percent of majors in the course had no effect.

The major problem in interpreting the studies that find a statistically significant positive association between course grades and students' evaluations of their instructor and between cognitive achievement and students' evaluations is sorting out cause and effect. If better instructors tend to emphasize the kind of higher learning that students are likely to protest (as costly to acquire), and if the standard tests of cognitive achievement fail to pick up these effects, the better instructors might have classes that do poorly on standardized tests and give low ratings.²¹ This, however, is speculation.

²¹ For a contrary view, namely, that student evaluations will encourage instructors to teach lower level cognitive material (e.g., basic facts and concepts with applications to narrowly defined problems, rather than analyzing, evaluating, and synthesizing broad real-world problems), see Michael Everett [45, 1977]. He argues that because student evaluations seem to depend mainly on clarity of communication in the classroom, well-prepared lectures, and instruc-

Other than cognitive achievement (perhaps in a negative direction) and grading standards, what do student course evaluations measure? In a careful study of 4,996 course evaluations from the Graduate School of Business at Stanford University, Lumsden found that specific characteristics like clarity of presentation, enthusiasm, and respect for student opinion had the largest positive impact on course evaluations [96, 1974]. Sensitivity to a student's needs, interest in the student as a person, and availability outside of class were unimportant (although one might expect different results from undergraduates). Costin, Greenough, and Menges, summarizing many studies of course evaluations, concluded that "such attributes as preparedness, clarity, and stimulation of students' intellectual curiosity were typically mentioned by students in describing their best instructors" [35, 1973, p. 52].

The only widely used student evaluation form is the Purdue Rating Scale for College Instructors. It is a questionnaire consisting of 28 items. Each student rates the instructor from A to F on five broad categories of questions—personal characteristics (*e.g.*, patience and understanding), objectivity (*e.g.*, willingness to listen to and talk about divergent ideas or viewpoints that oppose his/her own), exposition skills (*e.g.*, concise presentation), tests and grading (*e.g.*, impartiality), and subject matter knowledge.²² Most evaluations of instruction are made with the use of locally designed and administered questionnaires, making it difficult to compare the effects of instructional techniques on student assessments of teaching quality

across schools. The advantage of "home-made" course evaluation instruments is the greater faculty confidence in results of their own handiwork [60, Hansen and Kelley, 1973, p. 18].

A serious problem with empirical research on student evaluations of teaching involves the transformation of questionnaire data to numerical indexes appropriate for statistical analysis. The typical course evaluation form provides a scale of five alternatives, from much better to much worse, on which students are asked to rate the particular course or instructor. Thus the questionnaires ask for ordinal data. Then the data are usually coded on a one to five scale and employed in the analysis. Raymond Battalio, Joe Hulett, and John Kagel have shown that conclusions based on regressions using data obtained from an ordinal rating can theoretically be reversed by performing entirely legitimate order-preserving transformations on the measured variables [15, 1973]. The real question, however, is whether this difficulty is important in a practical sense [140, Siegfried, 1973]. Without knowledge of the actual underlying cardinal scale, it is impossible to evaluate the seriousness of the problem.

Student course evaluations have probably received so much attention because they are frequently thought to influence faculty behavior. McKenzie and Staaf [104, 1974] and Hansen and Kelley [60, 1973] have used microeconomic theory to demonstrate how faculty might respond to changes in the reward structure. Furthermore, the effect of changes in time constraints (say, by varying the course load) are ambiguous because income effects may swamp substitution effects. Becker has shown rigorously that increasing the weight assigned to high quality teaching in the faculty income determination process may be ineffective unless improvements are made simultaneously in the accuracy of measuring teaching ability

tor enthusiasm, rather than on stimulating intellectual curiosity or illustrating how economists analyze broad complicated problems, they may encourage the teaching of lower level material that can most conveniently be taught with those instructional skills.

²² For more detail on the Purdue Rating Scale, see H. H. Remmers and J. A. Weisbrodt [123, 1965].

[17, 1979]. In a further extension Cliff Huang demonstrates that the faculty work response to changing relative rewards to teaching and research may be ambiguous if the production processes for research and teaching are less than certain [68, 1979].

In an empirical study of the faculty salary determination process at the University of Wisconsin, Siegfried and Kenneth White found that teaching quality (measured by course evaluations) was rewarded positively [146, 1973]. The coefficient on a variable sensitive to the extremes in teaching performance was statistically significant at the .10 level. The importance of its size depends on the cost of improving teaching ratings. David Katz reported contrary findings about the rewards for teaching at a large midwestern university [75, 1973]. Katz, however, used a binary variable, which discriminated only between teachers above and below the median and consequently did not characterize the salary determination system as it was implied in his own interview reports, namely, that salaries are sensitive to extremes in teaching quality, but respond very little to differences close to the average. Siegfried and White [147, 1978] have shown that the differences in specification—sensitivity to extremes versus sensitivity only around the median—may cause the difference between Katz's findings and theirs. In a comprehensive analysis of the determinants of faculty salaries that was based on data gathered by the American Council on Education as part of a 1972–73 national cross-section of faculty, Howard Tuckman found that faculty who had won an award for outstanding teaching earned no more than others, *ceteris paribus* [167, 1976].

The Siegfried and White study examined salary differentials within an academic department; Katz's sample included faculty from various departments within a single university; Tuckman's data

base (for the result reported here) consisted of economists at various institutions. These findings can be reconciled if the recognition (and reward) for teaching is local. For research there are national standards and visible evidence of performance (publication), while teaching reputation is usually well known only to local colleagues. Thus salary differences based on teaching ability may appear only within individual departments.

C. Impact of Understanding and Attitudes on Behavior

Economists generally measure the value of education by the discounted present value of the expected difference in earnings streams (net of direct costs) attributable to it. There are at least two major reasons this approach has been conspicuously absent from the literature of economics education. (1) Economics is usually offered as part of a liberal arts education. Because it is only a small part of a large program of education, it is difficult to attribute earnings differentials to specific courses in the program. (2) Many economists believe there are benefits to economics education that individuals cannot appropriate at reasonable cost, in which case private earnings will not fully reflect the total benefits from economics education. This frequently used "citizenship argument" for economics education means that economics education raises students' sensitivity to the political, economic, and social system of which they are a part and increases the intelligence with which they participate in it (*e.g.*, by voting) [58, Hansen, 1977].

The difficulties in assessing public benefits to economics education have been summarized by McKenzie: "Before the public benefits . . . can be acquired through the political process, the student . . . must have sufficient incentive to maintain the human capital stock that he has acquired. . . . Public choice theory

predicts that the typical individual student-citizens will not have sufficient incentive to incur these costs. . . . Finally, economic education must overcome the tendency of people, in spite of what they know about the economic merits of legislation, to vote their own private interests" [103, 1977, p. 10].

One approach to assessing subsequent student behavior compares the fraction of students in an experimental introductory course who go on to major in economics with the fraction from a conventional introductory course. The notion is that students who have favorable attitudes toward economics and/or believe there is value in economic reasoning are more likely to major in economics and also more likely to capitalize on their economic reasoning skills in later life. Unfortunately this measure of effectiveness has several disadvantages: (1) The percentage of students from a class who go on to major depends on the denominator as well as the numerator, and many exogenous factors (such as curriculum requirements of other, non-economics, major departments) may influence the number of students taking a course in economics. (2) An introductory economics course that provides an accurate portrayal of what a student might expect in future courses in the major may (appropriately) dissuade many from majoring. But this does not mean the course is a failure, any more than it means that the students' talents and/or interests are "bad." A model of student curricular choice formulated by McKenzie and Staaf [104, 1974] and tested empirically by Alan Freiden and Staaf [53, 1973] implies that students who are superior in both verbal and mathematical ability tend to switch majors less frequently than other students because they do not have to compromise their interests in order to achieve the highest grades possible for them. Thus, if student enrollment in voluntary experimental courses is correlated with abilities,

inferences from the fraction who subsequently major in economics may be contaminated. (3) The advantage of effective experimental teaching methods may accrue to other courses or leisure time, depending on students' time allocation decisions. (4) When there are changes in the demand for graduates of different disciplines, the relative benefits from majoring in certain fields will be altered and induce a shift in student selection [52, Richard B. Freeman, 1971]. Since the introduction of new teaching technologies may accompany exogenous shifts in demand, there is risk of incorrectly attributing "success" to a new learning technology if majoring in economics is part of the output matrix.

D. Attitude Sophistication and Values

One of the goals of economics education is greater sophistication about policy issues. Economics training may also make students more conservative or more liberal. Researchers have attempted to measure both effects.

William R. Mann and Daniel R. Fusfeld have argued that "the attainment of a high level of attitude sophistication should be as much a goal as the proper manipulation of supply and demand schedules" [98, 1970, p. 125]. They developed a Questionnaire of Economic Attitudes (QEA) and concluded that the attitude sophistication of an experimental group improved while that of a control group without economics instruction did not. Mitchell P. Rothman and James H. Scott, Jr., criticized the measuring instrument as favoring liberals [126, 1973]. If they are right, Mann and Fusfeld's evidence suggests that economics makes students more liberal. Howard Tuckman [166, 1975] used a 20-item Attitude Sophistication (AS) test along the lines of Mann and Fusfeld's. He found that "an increase of about 3.3 points on the pre-AS exam adds one point to the post-AS score and 0.17 points to the final grade." He attributed the significant cor-

relation of the AS score with the grade to its measuring "economic reasoning rather than prior knowledge" [166, 1975, p. 36]. The AS test was not published. If it is free of bias, further research with it might be worth pursuing.

The 1973 article by Rothman and Scott [126] and another paper by the same authors [139, Scott and Rothman, 1975] report the results of administering a Social Opinion Questionnaire (SOQ) to students at Carnegie-Mellon. Whereas the QEA and the AS were intended to measure progress toward increased sophistication from studying economics, the SOQ was intended to "separate the students along a 'liberal-conservative' dimension." In the first study published, Rothman and Scott found that "conservative answers are almost invariably associated with higher expected *TUCE* scores at the start of the semester [*i.e.*, before the study of economics began]. Equally important is the fact that the most significant questions are those associated with an individual's preferences for the capitalist market system" [126, 1973, p. 121]. At the end of the course, which included some microeconomics but emphasized macro, there was no significant difference between liberals and conservatives on Part I of the *TUCE* (which, like the course, includes some micro but emphasizes macro). The authors concluded that the SOQ measures something other than economic knowledge, that conservatives on entering the introductory course know more economics than liberals, and that there is no systematic bias in the *TUCE*, Part I, when used at the end of the introductory course. Since the sample size was only 49 students at one school, these conclusions must be deemed exceedingly tentative.

The second paper throws more light on the effect of an introductory psychology course than on an introductory economics course. The psychology instructors were more liberal than their students and influ-

enced them in a liberal direction. Scott and Rothman report that "the college experience is a liberalizing one [139, 1975, p. 109], but the psychology course had a stronger liberalizing effect than the economics course [139, p. 110].

Kim Sosin and Campbell R. McConnell investigated the effect of an introductory macroeconomics course on attitudes toward income distribution [156, *c.* 1978]. They found a significant shift in attitude toward more egalitarianism compared to a control group. The largest shifts were by students with the least incoming tendency toward egalitarianism and those with the highest grades. Inasmuch as Milton Rokeach [125, 1973] has shown that in the United States the liberal-conservative spectrum can be reduced to preference for equality, the finding of Sosin and McConnell that a macro course made students more liberal is convincing even though their survey was confined to a single issue.

In conclusion, it has not yet been established that attitude sophistication is a measurable output. On the question of political orientation, the evidence cited tends toward the conclusion that economics has a liberalizing influence. But the hypothesis that *microeconomics* gives students an understanding of markets that makes them more conservative has not been disproved.²³

E. Distribution of Benefits

The economic efficiency of new teaching methods should be assessed by comparing marginal benefits with marginal costs. Marginal benefits depend on the impact of the method on direct output units (*e.g.*, cognitive achievement, student atti-

²³ Cf. Stigler [158, 1959] and Fels [50, 1978]. Some authors have concluded that microeconomics had a conservative influence but, having doubts about their evidence, we do not cite them. Fred A. Thompson found that increased knowledge of international economics was associated with a shift to a more favorable attitude toward free trade [163, 1973].

tudes) and the value that recipients place on the outputs. Hansen, Kelley, and Weisbrod observe that different opinions about who are "most 'important' to teach may be at the root of much of the controversy [surrounding instructional techniques]" [61, 1970, p. 365]. For example, students differ in previous preparation, verbal and mathematical abilities, career goals, sex, social concern, and many other characteristics. Consequently, they are likely to benefit differentially from any particular instructional method. It is impossible to assess the economic efficiency of a particular method without valuing the benefits received by various individuals. Typically we rely on market forces and equilibrium competitive prices to assign values to the marginal outputs. The peculiar financing of higher education, the market failures in appropriating benefits, and the inability to attribute consequences to causes (earnings to courses) make ascertaining such values difficult. Nevertheless, it is important to consider explicitly the distributional effects.

The two methods for incorporating distributional effects into typical production function studies of instructional methods are either to include interaction variables in the model—interactions between the experimental treatment variable and the important characteristics that vary among students, *e.g.*, sex, family income, natural ability—or to estimate the relationship separately for various groups of students. These techniques have been used occasionally in recent years (*e.g.*, Kelley [80, 1972]; Allison [2, 1976]) and the findings reveal the importance of considering separately the impact of instructional methods on various groups of students. However, little progress has been made (indeed, maybe none can be) in assessing the differential valuation that different students place on receiving increments of cognitive achievement or changes in attitudes.

III. The Impact of Human Capital and College Environment on Economics Education

This section summarizes the research findings on the effect of student human capital, faculty human capital, college environment (other capital), and student effort on learning economics. There is, to our knowledge, only one study of the effect of faculty effort on learning economics. Allison [2, 1976] found that instructor preparation time increased student achievement, *ceteris paribus*.

The focus of most evaluative studies in economics education is on new learning technologies. Capital and variable inputs are usually included in the regression equations in order to control for other factors that may be correlated with student learning and the presence of the innovative technique (*e.g.*, students with higher aptitudes may be more inclined to elect innovative techniques if they are given the option, in which case one might erroneously attribute their better performance to the innovation when, in fact, it was caused by their superior aptitude). The number of studies that consider such variables is enormous. Therefore we limit references to them.

A. Student Human Capital

Three types of variables have been used to assess the impact of student human capital on economics education output. First, a variety of general aptitude measures have been employed. Most studies find college entrance examination scores (*SAT*, *ACT*) to be positive and significantly associated with economics test performance. Verbal *SAT*'s seem to be more important than quantitative *SAT*'s for the *TUCE*.²⁴

²⁴ Mathematical aptitude seems to be more important in scoring well on locally constructed tests. This may reflect the emphasis of the *TUCE* on applications.

This finding is supported by the usual non-significance of previous mathematics courses in such regressions. High school class rank, a measure of general aptitude, usually has a positive impact on post-test examination scores in economics. Measures of student maturity, such as age, year in school, and number of previous college credit hours, usually show no relationship to cognitive performance.²⁵ The few studies that have included measures of students' socioeconomic backgrounds have found such variables as family income and/or parents' education to be unimportant.

Second, measures of prior knowledge of economics have produced mixed results. The effect of high school economics courses on performance in college economics has been widely investigated, but the results are inconclusive. The most recent study concludes that after adjusting for other factors, students who had taken previous economics courses did not begin their principles course with significantly more knowledge, nor did they learn significantly more during the semester [120, John Palmer, Geoffrey Carliner, and Thomas Romer, 1979]. The most obvious predictor of the post-test score in introductory economics is the pre-test score. It is almost always found to be positive and significant. However, when gap-closing measures are used, the pre-test score has been found to exert a negative influence on the post-test score. This suggests that students who know more economics at the beginning of the course learn relatively less during the course, but Becker and Salemi have shown that this apparent paradox may be explained by simultaneous equations bias in most of the studies [18, 1977].

Third, many studies have included stu-

dent major and indexes of student interest, presumably to test the hypothesis that learning is related to motivation. The results of these tests are mixed, which may reflect the variety of instruments used to measure interest.

B. Faculty Human Capital

Several studies have found that years of teaching experience, instructor *TUCE* scores, and instructor's graduate school grades are positively related to learning in introductory economics classes. In addition, the studies that have attempted to explain cognitive achievement with course evaluation scores may be directed toward the question of whether faculty human capital affects student learning. If in evaluations students rate highly the characteristics of instructors that help them learn more efficiently, then we would expect students of professors with higher evaluations to perform better on examinations. Many studies that use multiple instructors of a single course find that the particular instructor does make a difference. Apparently differences among instructors are not entirely captured in the models.

C. College Environment

Studies of class size are almost unanimous in finding no influence on test scores. Harry Levin found no difference in cognitive achievement between students in classes of 30 and students in classes of 80 to 120 [90, 1967]. However, students rated the course higher if they were in smaller classes. The latter finding has been confirmed by Mirus [106, 1973]. Lewis and Dahl, while finding the usual lack of impact of class size on *TUCE* performance, discovered a negative effect of class size on students' critical thinking skills [92, 1972]. In general, it appears that students are happier and perhaps learn to think better in smaller classes, but performance

²⁵ The non-significance of many control variables may be due to multicollinearity, in which case a test of their joint significance would be more appropriate.

on standardized tests is independent of class size.

Attiyeh, Bach, and Lumsden found that students from larger colleges and from colleges with higher schoolwide SAT averages did better on the *TUCE* [6, 1969]. In addition, they found substantial variation in performance on the basis of type of school.

Most research has found that two semesters of introductory economics yield greater understanding than one. The latest of these studies [42, Ralph Elliott, M. Edwin Ireland, and Teresa Cannon, 1978] reports that two-semester students at Clemson University do significantly better on each of various functional forms of *TUCE* scores. The gains, however, are small in comparison to a one hundred percent increase in inputs (a second semester).

In most studies of the question, the choice of textbook did not matter. In Saunders's study there was no significant difference in *TUCE* scores for students using any of five leading introductory textbooks listed by author's name [132, 1973, p. 60 and Table 15]. For seniors who had had the introductory course two years earlier, however, those who had had textbooks other than the five named did significantly worse, as did those alumni whose textbook was unknown. Attiyeh, Bach, and Lumsden, on the other hand, found that students using either of two leading conventional textbooks did significantly better than those using any of ten books lumped together in an "all other" category, whereas those with a third leading text did not [6, 1969, pp. 220-21]. As pointed out in *Subsection IV.D* below, they also found a difference between two programmed texts.

D. Student Effort

It is not clear whether student effort should be an independent (control) or de-

pendent variable. Indeed, one of the important insights of the application of microeconomic theory to the learning process is that students may well choose to take efficiency gains in the form of reduced time inputs into their economics course [104, McKenzie and Staaf, 1974].

A number of studies have attempted to correlate student effort with test performance, the best of which are by Becker and Salemi [18, 1977] and Allison [3, 1977]. With few exceptions, these studies have found no impact of study time on performance. In a simultaneous equations model, Allison finds low positive elasticities of achievement with respect to student effort [3, 1977]. However, her results indicate that students "learn to learn"; the elasticity, while remaining low relative to that of student ability, quadrupled from the first to the second semester course. Other measures of student effort confirm this conclusion. Attendance and student course load (which measures competing demands on student time) do not appear related to performance on standardized tests.²⁰

Grade point averages are sometimes interpreted as measures of student effort, especially when included simultaneously with aptitude test scores. Grade point average (GPA) is usually found to be positively related to test performance, which might indicate that GPA is a better measure of aptitude than student effort.

In sum, it appears that a student's general (especially verbal) aptitude is the most important determinant of learning. Socioeconomic background, prior economics courses, mathematics preparation, class size, textbooks, and study effort do not seem to matter very much. The evidence on the effect of student interest on test performance is mixed.

²⁰ Class attendance does seem to be important for achievement on locally constructed tests, but not standardized ones [119, Donald W. Paden and M. Eugene Moyer, 1969].

IV. *The Impact of
Alternative Teaching Methods
on Economics Education*

A. *Games and Computer-Assisted
Instruction*

The computer has been employed in economics instruction in two general ways: (1) games, simulation models, and demonstration routines (CAI); and (2) study management systems (CMI). CAI includes the popular macroeconomic models of the economy, in which students attempt to set policy parameters in order to achieve a menu of specified macroeconomic goals that involve trade-offs. CMI consists mainly of review routines (short quizzes) with instant feedback to students and individualized instructions to students wanting to know the most efficient study strategy to pursue. It is discussed in the next section.

Because of their sequential nature, computational requirements, and record-keeping needs, many games and simulations utilize computer facilities, especially time-sharing. Games, however, can be played without the use of computers. Most games involve role playing—a stock-market trader, a manager of a firm under various degrees of competition, an urban planner, or an advisor to the finance minister—where the student is required to map out strategies over time. Thus games tend to emphasize the interactions in an economic situation and are more likely to teach analytical methods successfully than recognition and understanding of economic terms and concepts.

In a 1974 review of the computer-assisted instruction (CAI) literature, Soper listed 45 published and unpublished reports of the use of CAI in teaching economics [152, 1974]. Most of the reports described the use of games and simulations in teaching principles of economics or intermediate macroeconomics. Since

almost all of the games that have been evaluated use the computer, we henceforth use CAI to mean games and computer-assisted instruction.

There are several reasons to expect CAI to influence learning. Instant feedback (at least with time-sharing), novelty, and convenience all should serve to alter students' learning functions. Quick reinforcement of correct responses or immediate explanation of incorrect responses has long been one of the most widely accepted principles for improving learning. The novelty of computer-assisted instruction may improve student attitudes and, in turn, learning. The convenience of time-sharing may influence learning by permitting students to use their time more efficiently.

Simulation models, which apply the principles gleaned from comparative statics to a dynamic world, may give students an appreciation for the difficult problems confronting policy-makers. Computer oligopoly games instill respect for the dynamic strategies that complicate the real industrial world and require complicated models to characterize reality. Games are (usually) less abstract than lectures and may stimulate learning because they are realistic. Finally, most students seem to like games, and student attitudes may affect cognitive performance [23, Samuel Bowles, 1970].

The experiments testing the effectiveness of games and/or computer-assisted instruction assess the impact of CAI on (1) cognitive evaluation instruments, such as the *TUCE*; (2) student attitudes; (3) the lasting effects of greater cognitive achievement of students who experienced CAI; and (4) the distribution of benefits between high- and low-achieving students. Overall, the conclusions about the effectiveness of CAI in improving understanding of economics are pessimistic.

Wentworth and Lewis carefully evaluated the noncomputer learning game

called Marketplace at two Minnesota junior colleges during 1971-72 using multiple linear regression analysis and a sample of 149 students divided between two instructors and users and nonusers of the game (four cells) [176, 1975]. They found that playing the game as a substitute for eight class periods of conventional lecture instruction has a statistically significant impact on students' gain in *TUCE* scores during the semester. The effect, however, was negative. Students attending the lectures gained 1.43 points more on the *TUCE* during the semester after controlling for intelligence (*ACT*), age, college, high school economics background, sex, and student interest in economics.

The most complete study of CAI is reported by Davisson and Bonello [40, 1976]. They describe the development of computer-assisted instruction in principles of economics at Notre Dame during the 1973-75 period and report the results of an evaluation of its cost effectiveness. Review routines comprise the major part of the CAI program at Notre Dame and consist of multiple choice questions with prompts and verification statements. There are also demonstration and game simulation routines. The impact of CAI on cognitive performance in the Notre Dame experiment was evaluated using the *TUCE*. The findings were similar to many of the other experiments with CAI—no significant difference between the experimental and control groups.

John Chizmar *et al.* [31, 1977] adapted Alan Blinder's [19, 1973] methodology for separating shifts in the production function from shifts in the coefficients of variables included in the model and found that Illinois State students using the Notre Dame CAI package performed slightly better on the *TUCE*; but they did this *in spite of CAI*, primarily because of greater ability.

In a regression analysis of the effectiveness of CAI games in a macroeconomics

course at Arizona State University, Steven Cox found no effect of CAI games on student performance on multiple choice tests [36, 1974]. He experimented with nonlinear relationships between the independent and dependent variables, but found that a simple linear model produced the same conclusions as the more complex specifications.

In an evaluation of the effectiveness of one micro and two macro games at St. Olaf College, E. David Emery and Thomas Enger assessed student performance on various *types* of multiple choice questions [43, 1972]. They found that CAI led to higher achievement on questions requiring analysis and policy decisions.²⁷ However, in a follow-up to the St. Olaf CAI experiment, Emery and Jean Schoene found that the initial advantage of CAI students had completely disappeared over 16 months [44, c. 1974].

Most of the reports of CAI activities indicate that students enjoy playing computer games. However, Davisson and Bonello's extensive analysis of the impact of CAI on student attitudes revealed no differences between the CAI users and control groups [40, 1976].

On the question of the distribution of benefits from CAI, Cox found that all students except those who earned C grades benefited [36, 1974]. Emery and Enger found no difference in benefits on the basis of student aptitude [43, 1972]. Bonello, Davisson, and Swartz report that lower achievement students in the Notre Dame experiment benefited from CAI, while better students did not [20, 1978]. This occurred in the presence of a nonsignificant overall effect and illustrates why it is important to identify precisely the tar-

²⁷ Because students in the experimental group did not do significantly better on all types of questions, they concluded that their better performance on some questions was not due to an incentive (Hawthorne) effect. There is no reason to expect incentive effects to influence only one type of learning.

get group and assess the impact of educational experiments on specific subgroups of students.

Davisson and Bonello attempted to assess the costs of implementing CAI [40, 1976]. They estimated that start-up costs were \$20,000 in 1973-74, but their estimate is based on a very low valuation of faculty time (\$1,500 per month, implying a \$13,500 academic-year salary, which in 1973-74 was about the starting salary of assistant professors). The faculty who have the skills to develop CAI are also likely to be the faculty who have skills sufficient to insure a high opportunity cost of time during summers.

Operating costs can be assessed on either a budgetary or economic basis. Budgetarily, CAI will probably cost much more than conventional instruction because universities tend to assess user charges for computer use but not for the alternatives such as library use. However, in real economic costs, the difference may be much less. An input that can be dispensed with if CAI is adopted is graduate teaching assistants. Davisson and Bonello found that CAI was no more expensive than graduate teaching assistants meeting discussion sections once a week. On the other hand, if the opportunity cost of graduate TA's is very low to a department (that is, they are there and supported financially, not to provide instruction to undergraduates but because the faculty wants a graduate program for some other reason), then CAI will be more expensive.

Overall, games and CAI in economics do not appear to be the route to nirvana they were once expected to be. CAI appears to generate no more (or no less) cognitive achievement,²⁸ but probably costs

more than conventional pedagogical methods. In addition to instructor and computer-facility costs, most games and computer-simulation exercises are student-time intensive. They are unlikely to be generating benefits in the form of released student time.²⁹

B. Computerized Study Management

There have been at least three large-scale efforts to utilize computer facilities to individualize and improve student study management. Kelley [78, 1968; 80, 1972; 82, 1973] reports on the effectiveness of the Teaching Information Processing System (TIPS) at the University of Wisconsin, Bernard Booms and D. Lynne Kaltreider [21, 1974] on Computer Generated Repeatable Testing (CGRT) at Pennsylvania State, and Donald Paden, Bruce Dalgaard, and Michael Barr [118, 1977] on the computer study management system (PLATO) at the University of Illinois. These three experiments have been conducted at large state institutions, where large classes and impersonal atmospheres are more likely to create need for individualization of instruction.

Computerized study management systems usually administer periodic short quizzes and then provide rapid feedback to students. Kelley's system evaluates students' learning problems (from "surveys" that do not count toward students' grades) and provides a personalized assignment suited to the level of understanding of each student. In addition, such programs usually provide record-keeping and test analysis and grading services.

Study management systems are expected to improve student learning by "directing student activity." This implies that someone other than the student

²⁸ One possible explanation for the negative findings can be found in a warning from one of the pioneers in developing computer games, Myron Joseph, who cautioned that "there is some danger that the complexity permitted by the computer will obscure the learning objectives of simulation" [73, 1970, p. 95].

²⁹ Short, non-computerized classroom games, however, may be very inexpensive. Unfortunately there is an absence of solid controlled assessments of such games, so little can be said about their efficiency.

knows best how the student learns most efficiently. Study management systems could influence learning by improving student attitudes or by providing individualized help. In addition, Kelley uses the TIPS system to inform the lecturer and teaching assistants of student progress. The study management systems also encourage students to keep up, thereby avoiding the deleterious effects of cramming.

Kelley performed controlled experiments of the effectiveness of TIPS and found it had a significant positive effect on achievement [79, 1970; 80, 1972]. The effectiveness of TIPS is implemented by providing additional, less difficult assignments to low-achieving students, and fewer, more difficult assignments to high-achieving students, thereby directing teaching and student resources to the place where their marginal product is higher. Kelley reports that TIPS is not more expensive than conventional instruction because it may cut back on assignments to bright students while increasing assignments to low achievers [80, 1972, p. 426]. A follow-up test a year later indicated that the TIPS cognitive differential was maintained [80, p. 426]. Most of the benefits of TIPS appear to accrue to low achievers [80, p. 425].

The Pennsylvania State system (CGRT) is similar in many respects to TIPS, but the computer tests count in students' grades, while Kelley uses them only for study practice. Booms and Kaltreider found higher mean *TUCE* scores for CGRT students than for a control group [21, 1974]. Student opinions of CGRT were favorable, but the system cost almost 50 percent more than conventional instruction.

PLATO consists of instructional sequences and elaborate testing and record-keeping capabilities. The Paden *et al.* evaluation of PLATO found that mean scores on an unspecified cognitive test

were higher, but they did not attempt to hold human capital and utilization rates constant [118, 1977]. This deficiency is dangerous in view of the findings of Chizmar *et al.* that students who elect computer-assisted instruction may have characteristics that permit them to do well in spite of CAI [31, 1977].

In sum, the computer-aided study management systems seem to perform much better than games and simulation routines. Although there were no clear-cut cost advantages to computerized study management, it does not appear to be significantly more expensive.

C. Programmed Instruction

Programmed instruction, brainchild of the behavioral psychologist B. F. Skinner (see Skinner and J. G. Holland [148, 1961]) has been the subject of one major and a number of minor studies in economics education. The major study, by Attiyeh, Bach, and Lumsden (A-B-L) [6, 1969; 7, 1970], found a large gain in efficiency from using programmed instruction. The minor studies have not seriously challenged this result.³⁰ Nevertheless, programmed instruction is not widely used in economics.

Programmed instruction is based on certain psychological principles of learning, particularly positive reinforcement,

³⁰ In the chief of the minor studies, Attiyeh and Lumsden showed that programmed instruction written for high school use was superior to the alternative in a Palo Alto school [8, 1965]. Bach found that a superior teacher could outperform the better of the two programmed texts used in the study by Attiyeh, Bach, and Lumsden [6, 1969]. Soper [150, 1973] and Donald Darnton [39, 1971] have developed methods of overcoming student antipathy to programmed instruction. Other studies include Stephen Buckles and Marshall E. McMahon [27, 1971], Paden and Moyer [119, 1969], Fels and Dennis R. Starleaf [51, 1963], Fufeld and Gregory Jump [54, 1966], Thomas Havrilesky [66, 1971], Dennis Weidenaar [169, 1972], Attiyeh and Lumsden [10, 1972], Lumsden [95, 1967], Phillip W. Tiemann, Paden, and Charles J. McIntyre [164, 1966], and William C. O'Connor [117, 1974].

active involvement of the student, prompt feedback, encouragement through psychological rewards, and instruction of complex ideas by breaking them down into small increments. The material is presented in a succession of "frames," each usually consisting of one or two sentences with blanks for the student to fill in. The students find out immediately whether their answers are correct by looking at the bottom of the page or on the next page. They almost always *are* correct. They therefore are supposed to get a feeling of accomplishment that will encourage them to go on, and on the infrequent occasions that they are wrong, their errors are corrected immediately. Programmed instruction contains a great deal of irregular repetition to reinforce learning of important concepts. In contrast to the usual encyclopedic textbook, the writing of programmed instruction requires the authors to decide exactly what they want to teach and to omit everything else, a characteristic designed to make learning more efficient.

A-B-L conducted a nationwide experiment on programmed instruction involving 48 schools and 4,121 students. Each participating school established three test groups. Group I used one of two programmed texts exclusively for three weeks, studying a total of 12 hours per student on the average. At the end of this period, the students were tested (and from then on received conventional instruction). Group II students received conventional instruction supplemented with the programmed learning text (on which they spent an average of eight hours). Students in Group III had only the conventionally taught course. Groups II and III were tested after they had had an average of seven weeks of their respective courses.

A-B-L used regression analysis with the number of correct answers on a preliminary version of the *TUCE* as the dependent variable. All the variables rep-

resenting student characteristics (*e.g.*, educational level, sex, SAT score) were statistically significant at the 95 percent confidence level and quantitatively important. SAT score was the most important single determinant. In terms of school characteristics, average freshman entrance examination score, school size (with a positive coefficient), and school type were statistically significant. The coefficient for state colleges was positive and statistically significant, the coefficients for liberal arts colleges and for large state universities were not significantly different from zero, and the coefficient for "prestige" schools was significant and negative.

The major results were: (1) on the average, spending 12 hours studying a programmed text over a three-week period was approximately equivalent to spending seven weeks in a conventionally taught course; (2) students using programmed learning performed relatively better on applications than on recognition-and-understanding questions; and (3) students had a positive attitude toward programmed learning, generally considering it more effective than interesting.

The A-B-L study is notable for being one of the few in economics education based on a large sample of schools and students. It is also notable for having been largely ignored by a profession that normally takes a great interest in questions of efficiency. Despite the finding that programmed instruction can accomplish almost as much in three weeks as conventional instruction in seven, there was no rush by economics professors to adopt it.

D. Personalized System of Instruction

The personalized system of instruction (PSI) pioneered by the psychologist Fred S. Keller [77, 1974] is quite different from Skinner's programmed instruction but is based on similar principles. The enthusiasm it has generated in fields other than

economics, though reminiscent of the early days of teaching machines, appears to go deeper and have more staying power.

Like programmed learning, PSI requires instructors to decide exactly what they want students to learn. Students are given short assignments accompanied by specific statements of behavioral objectives.³¹ Students study the assignments at their own pace. When they feel ready on the first assignment, they take a test which is immediately graded for mastery (usually defined as 90 percent). A student who passes goes on to the next assignment. Students who do not pass are recycled—*i.e.*, they study the assignment some more and take another test, repeating the procedure as often as necessary to pass. The grade in the course depends on the number of assignments completed. Proctors (usually undergraduates who have already had the course) are available to help the students and to administer and grade the tests. There are few, if any, lectures. As with programmed instruction, feedback is immediate, and students are expected to get a sense of satisfaction from achieving mastery (and be motivated by it to go on). PSI has been used in many fields and schools. In the vast literature on PSI, controlled experiments typically show that students like PSI and think they learn more from it than from conventional instruction. Objective data show that they have in fact learned at least as much, sometimes more. They work harder but feel the extra effort is worth it.

Fels's PSI course [49, 1974] (see also Evans Jenkins [69, 1977]), in contrast to the one at Harvard to be described next, followed the Keller model closely. Although

the results of an elaborate evaluation were similar to those of PSI courses generally, there was no objective confirmation of the students' impression that they had learned more; contrary to the usual experience, the students worked no harder than in a conventional course; and the chief educational gain went to the proctors, who apparently learned a good deal more than they would have from taking intermediate economics courses. (See Siegfried and Stephen H. Strand [143, 1976] and Siegfried [141, 1977].)

Self-pacing is a fundamental part of the PSI method. Quick students are not held back by slow ones and may finish the course early. Slow students are not compelled to stay in lockstep with the class. All students can work at times most convenient for their other activities.

Self-pacing can be adopted without going the whole PSI route. In the fall of 1972, Harvard began an experiment in which self-paced instruction was used over a three-year period in three sections each with 25 to 30 randomly-selected students (Allison [2, 1976; 3, 1977]). The distinguishing features of the program were: (1) the course was divided into only eight units per semester for separate testing (usually PSI courses are broken into 20 or 30 units); (2) the sections had regular class meetings with voluntary attendance; and (3) upper-classmen and graduate students were used as graders.

For evaluation it was assumed that students allocate effort among courses and other activities to maximize some utility function. A three-equation model was constructed with an effort equation, an achievement equation, and an enjoyment equation; later a fourth equation was added for major. These simultaneous equations were designed for estimation of a production function.

The self-paced effect on student enjoyment was found through repeated questionnaires to be largely attributable to a

³¹ That is, the students are told what they are expected to be able to do in operationally meaningful terms, such as: given the input requirements in man-hours for two countries and two commodities, determine the comparative advantage of each. "Understand the model of comparative advantage" won't do; it is too vague.

Hawthorne effect.³³ Self-pacing increased performance on tests by 15 percent with the differential increasing two years after the course. The advantage was greater for freshmen and for students with SAT scores below 700 than for more advanced and better students, but the differences were small and variable. Controlling for effort showed that self-pacing actually increased students' learning instead of merely inducing them to work harder. The results are consistent with the hypothesis that the crucial feature of self-pacing is the interaction between grader and student, a directed form of one-to-one instruction.³³

The positive results from self-pacing obtained at Harvard, together with the large number of successful experiments in other fields, suggest that self-pacing has a significant contribution to make to economics education.

E. Video

A televised course on "The American Economy" with John R. Coleman as the principal teacher ran on 182 CBS stations, 54 educational stations, and 5 independent stations during 1962-63. Of the 160 half-hour lessons, 128 were on economic content, the other 32 on teaching methods. The total audience was over a million, with 5,000 taking the course for credit at 361 colleges and universities. The results were evaluated in three separate research projects.

In a study reported by Saunders, the *TEU* was given to 71 television students, 113 Carnegie-Tech students taking a one-year course in economics, and a control group of 73 school teachers "substantially

identical to the TV students, but who had not watched "The American Economy" . . ." [127, 1964, p. 398]. The TV group scored approximately the same (40.9 questions right out of 50 with a standard deviation of 4.8) as the Carnegie-Tech sophomores (40.8) and significantly higher than the control group (33.9). A multiple regression analysis identified one other variable besides taking the TV course for credit that significantly affected the outcome, namely, a previous course in economics.³⁴

Campbell R. McConnell and John R. Felton reported a controlled experiment consisting of 27 students who took the CBS-TV course for credit matched with 27 students from a large live lecture course at the University of Nebraska who were similar in grade average, number of college hours completed, and course of study [100, 1964]. On the *TEU*, there was no significant difference between the two groups, but on 120 multiple-choice questions prepared by the authors, the live groups did significantly better on the 60 conceptual and 40 problem-analytical questions. Consistent with the *TEU* findings, there was no difference between the groups on the 20 factual questions. Since the TV course did not aim to go much beyond what the Committee for Economic Development's Task Force Report [34, 1961] specified high school students should know (omitting such concepts as marginal cost and marginal revenue), the TV course was evidently successful in its aims even though less successful in achieving the aims of the introductory college course. This study is a useful complement to the one by Saunders. The authors con-

³³ *I.e.*, students like to take part in experiments. This advantage of self-pacing would be lost once the method became routine.

³⁴ In another report Allison surveyed self-paced introductory courses at seven schools [1, 1975]. She concluded that the reports on them were more useful as descriptions of the forms and functions of self-paced courses than as evidence of their educational benefit.

³⁴ The value of the study is limited not only because it was confined to students and teachers in the Pittsburgh area but also because it used the *TEU*, a test designed for high school students. Nevertheless, the substantial correlation found between *TEU* scores and grades based on essay examinations (.70 for Carnegie-Tech students, .82 for the TV group) is reassuring.

firmed Saunders's findings that the TV students scored just as high on the *TEU* as students in a conventional course, but on questions designed for college rather than high school students, those in the regular course did better.

The most ambitious study of the TV course was a survey by the National Opinion Research Center reported by Bach and Saunders [13, 1965]. About 20 percent of the 65,000 high school social studies teachers watched the TV course at one time or another, but only 5 percent watched once a week or more. Over 5,000 students took the course for credit with 4,400 completing it successfully, 1,800 of whom were school teachers. A 25-item version of the *TEU* was given to 3,966 teachers. In a multiple regression analysis (in which R^2 was only .15), watching the TV course was by far the most important variable, its coefficient being twice as large as that for taking five or more college courses in economics. One or two previous courses made no significant difference, presumably because of forgetting. Taking the TV course for credit, as distinct from only watching it, had a negative coefficient, apparently because the first group was interested in credit, the second in learning.

Although the TV films were subsequently made available for educational use, "The American Economy" was essentially a one-shot affair, enormously expensive, apparently well worth the dollars spent, but not to be repeated for a long time. There is nothing in the research results on it to indicate that TV lectures are superior to live lectures, but they do suggest that TV can be just as good.

Although closed-circuit TV is now widely used to teach elementary economics in schools with large enrollments, published evaluations are scarce. An experiment in 1964-65 reported by McConnell used a matched-pair technique to compare teaching by McConnell on television;

McConnell in a large, live, lecture class; McConnell in a small, live, lecture class; and graduate assistants at the University of Nebraska [99, 1968]. In terms of teaching as measured by 170 multiple-choice questions, the differences were not statistically significant. In terms of student attitude toward teaching method, TV was significantly worse than each of the other options. McConnell stated that TV was cheaper but gave no data. McConnell and Charles Lamphear reported a follow-up experiment at Nebraska in which 440 principles students were given the option of a televised lecture course or a course with no lectures at all [101, 1969]. An Omnibus Personality Inventory test indicated no significant difference between the 354 choosing to view the lectures and the 86 choosing the alternative. The TV lectures were textbook oriented. For evaluation, a battery of multiple-choice questions was used, including 170 internally generated items, the *TEU*, and items provided by the committee developing the *TUCE*. There were no statistically significant differences at the 5 percent level. Attitude surveys revealed a significant preference for the lectureless method. A subsequent report by Lamphear and McConnell indicated that the lectureless group and the TV group did better than all students taught by graduate teaching assistants the previous year [87, 1970].

If true, the Nebraska findings would be very interesting, but it is not clear how much credence can be put in them. Taken at face value, they suggest that textbook-oriented lectures, whether live or on TV, have a value-added close to zero. But they are based on one instructor at one university.³⁵ They are, however, consistent with the findings of Attiyeh, Bach,

³⁵ There are other questions that might be raised about the Nebraska studies. The groups were not selected at random; the Hawthorne effect could account for some of the results; and the statistical methods are not highly sophisticated.

and Lumsden on programmed instruction [6, 1969]. They are also consistent with the hypothesis that different students learn in different ways.

Paden and Moyer found no differences at the University of Illinois in amount learned between groups taught by live lectures, TV, and programmed instruction, but the attitude toward live instruction was more favorable than toward the other two [119, 1969].

To conclude, the hypotheses that students learn as much from TV lectures but prefer live lectures are plausible, but sophisticated testing of them in economics has been limited.³⁶

F. Specification of Instructional Objectives

The premise that course objectives should be stated in terms of specific student behavior has intuitive plausibility. If instructors decide in advance exactly what they are trying to accomplish, they may improve their chances of achieving it, and if the students are informed of the goals, the efficiency of their efforts may be increased.³⁷ Ideally the instructor specifies observable behavior on the part of the student that represents not only the goal but also the means for telling whether the goal has been achieved. In the words of Saunders: "To be complete, an instructional objective should also contain a statement of the *conditions* in which the student should be able to do it, and a statement of the *criteria* that will be used to judge how well it is done" [134, 1978,

³⁶ Effective use can be made of video in teacher-training programs, which are discussed in a later section of this survey.

³⁷ A reader of the first draft of this article argued that stating objectives deprives students of training in distinguishing important from unimportant material. It may be replied: if that is the objective, the students should be so informed (otherwise distinguishing the important from the unimportant degenerates into "psyching the professor out") and the course should be designed to achieve it.

p. 68, Saunders's emphasis]. For example, "Given a price index with one base year indicated, the student will correctly convert the index numbers to those of another base year when the new base year is specified" [134, p. 72]. Saunders then gives three multiple choice questions illustrating how to measure accomplishment of this objective.

The goals of a liberal arts education cannot, of course, be reduced to a series of specific behavioral objectives. There is danger that concentration on those tasks that lend themselves to such specification may divert attention from other important purposes. (See Yates [178, 1978]. Cf. Section I.B. above.) Whether specifying behavioral objectives is useful is a testable hypothesis. The evidence from three studies is inconclusive. Dennis L. Nelson [114, 1970] found positive value from specifically stated behavioral objectives; James Phillips's study [121, 1972] was inconclusive; and Cheryl A. Casper [30, 1977] had negative results.

G. Graduate Student Instructors

Graduate students are used as teachers and teachers' aides in most major universities in the United States. GSI's may be less effective instructors because of their inexperience in classroom teaching, competing demands for their time that take priority over their teaching responsibilities (*i.e.*, their own study for the Ph.D.), ignorance of effective teaching techniques (which may be related to their inexperience), and weaker understanding of the substantive material covered in the classroom. On the other hand, GSI's may compensate for their lack of experience by their enthusiasm, their efforts to identify what the students do not understand, their approachability *vis-à-vis* regular faculty, and the greater rapport they frequently develop with their class [166, H. Tuckman, 1975].

There have been published evaluations

of the relative effectiveness of GSI's in economics at six institutions—Princeton, Hebrew University, Carnegie-Mellon, Indiana, Florida State, and Nebraska. The six studies all employ the same basic methodology: each compares the performance of students taught by GSI's with students taught by regular faculty, controlling for other differences that may affect relative student performance.

Wallace Oates and Richard Quandt compared the performance of students in classes taught by advanced graduate students with that of students in classes taught by regular Princeton faculty for eight semesters, 1965–69 [116, 1970]. The GSI's at Princeton had all completed their comprehensive examinations for the Ph.D. and attended a weekly meeting of all instructors in the course, at which time the subject matter for the coming week, various techniques for presenting this material, and any particular problems that had been encountered were discussed. The sample consisted of 2,336 students, about two-thirds of whom were taught by GSI's. Since the basis for comparison was performance on a common final examination that varied from semester to semester, comparisons were made separately for each semester. Oates and Quandt found that students of GSI's did better in one semester, students of regular faculty did better in two semesters, and there was no statistically meaningful difference in the other five semesters. Regression analysis to explain relative student performance each semester, controlling for grade average in other courses and SAT scores, confirmed the conclusion of no systematic difference in the performance of students of GSI's and students of regular faculty.

Among the GSI's there was enough variation in previous teaching experience to determine that those students of GSI's who had more experience did better than rookie instructors. Since this finding—that

experience counts—is confirmed in other studies, it appears that there are balancing effects at work: (1) the lack of experience of GSI's hinders the performance of their students, while (2) the enthusiasm, approachability, interest, *etc.*, of GSI's helps their students, so that the net difference between GSI's and regular faculty is negligible. The Princeton experience may not be generalizable to the wide variety of GSI teaching experiences in the United States due to the rather select group of GSI's (very advanced students) who teach at Princeton.

Morawetz compared the performance of 1,930 undergraduate students in 66 different classes at the Hebrew University from 1967–75 (excluding the war year of 1973–74) [108, 1977]. Thirty-five of the classes were taught by 12 different faculty members and 31 were taught by 16 different GSI's. These GSI's had all completed at least two years of graduate work in economics but received little guidance in teaching. In six of the seven years there was no statistically meaningful difference in grades on a common final examination at even the 10 percent significance level between students taught by faculty and students taught by GSI's.²² In the seventh year the students taught by faculty scored higher, but a regression analysis to control for student aptitude demonstrated that this effect was due to the superior students in regular faculty classes that year. In a comparison *within* the GSI group, Morawetz found little difference in student performance based on the level of experience of the GSI's [108, 1977].

Saunders compared the performance of 2,136 students of regular faculty and GSI's on the *TEU* and the *TUCE* at Carnegie-Mellon University (CMU) during 1964–69 [130, 1971]. Sixty-three classes were

²² A surprising finding was that the performance of students in classes taught by faculty had a higher variance than the performance of students in classes taught by GSI's.

staffed by faculty and 28 were staffed by graduate students. At CMU the instructors were given some advice on teaching, but substantially less than Princeton was providing. Difference-between-means tests, supported by multiple regressions controlling for student attributes, indicated that the students of faculty and the students of GSI's performed comparably.

The CMU results were confirmed, with minor exceptions, for thirty GSI's teaching about half of the 8,895 students in two introductory economics courses at Indiana University over eight terms, 1971-75 [133, Saunders, 1975]. Holding other things constant, students of the GSI's performed as well as students of regular faculty in the first semester course and did significantly better in the second semester course. Since most of the GSI's started teaching in the first semester course and then moved to the second semester course, the Indiana results are consistent with the hypothesis that teaching ability increases with experience.

Howard Tuckman compared the teaching of five faculty and three GSI's in twelve macroeconomics courses at Florida State during 1972-74 [166, 1975]. There were 548 students in the study. Tuckman used multiple measures of performance because GSI's and faculty may differ in teaching skills. Using regression analysis, he found that years of teaching experience was statistically significant in explaining differences in both cognitive achievement and student attitudes (in the expected directions), but the effects were small. Inserting a binary variable into the regressions for faculty status, Tuckman found no difference between students of faculty and GSI's in either cognitive or attitude dimensions. However, since faculty status and years of teaching are highly correlated, the difference between GSI's and faculty can be assessed only by observing the combined effect of years of experience and faculty status. When experience was

removed from the regression model (its effect presumably then captured by the faculty status variable), the results indicated that students of GSI's learn less but have more favorable attitudes. This difference seems to be a function of GSI's lack of experience rather than their status as graduate students.

McConnell [99, 1968] and Lamphear and McConnell [87, 1970] compared the performance of students of GSI's with students taught in a large televised classroom section; some live classes taught by McConnell, who has written a highly successful textbook and done considerable research in economics education; and a class taught without lectures at all. The *TEU* showed no difference between students of GSI's and faculty members; a test of 90 questions drawn from the *TUCE* testbank revealed that GSI's students performed worse than the control groups. This study did not control for teaching experience or attributes of the students in the classes. Although McConnell and Lamphear concluded that the GSI's students did worse than the students in the alternative classes, this finding is true for only one of the two test instruments used. If the finding that the students of GSI's at Nebraska did worse than the students of alternative instructional techniques is not explained by better than average instruction in the control group, it might be because GSI's at Nebraska in the 1960's were not required to have passed their Ph.D. comprehensive examinations and were given virtually no advice and training in teaching.³⁹

Evaluations of the effectiveness of substituting GSI's for regular faculty in teaching (mostly principles of) economics reveal that: (1) "While there are no doubt better and worse teachers, they do not divide themselves neatly into two groups with

³⁹ In the 1970's, however, Nebraska has been one of the dozen schools pioneering such training in teaching economics.

the labels of faculty and graduate students" [116, Oates and Quandt, 1970, p. 138]. (2) Instructor experience seems to improve the cognitive performance of students. Coupled with the findings of no difference in performance between the students of faculty and GSI's, this suggests that GSI's have compensating attributes that balance their lack of experience. (3) The results are ambiguous on whether undergraduate students like GSI's more or less than faculty. (4) More experienced GSI's (and perhaps better trained GSI's, if experience is interpreted as a method of self-training) seem to be better instructors. Those schools which offered special training to GSI's on their teaching and/or required more advanced students to teach (in which case they had observed more teaching themselves) seemed to have greater success with GSI's.

A direct test of the effectiveness of teacher training for GSI's was conducted by Lewis and Charles Orvis [94, 1973]. During the fall of 1971, seven GSI's taught 3 of the 4 weekly classes of 14 sections of principles of economics at the University of Minnesota. The GSI's were given no help in instruction. Students were pre- and post-tested using the *TUCE*, and data on student characteristics and attitudes were collected for all students. The same seven instructors were then used in the winter 1972 quarter to teach another 14 classes. During this quarter the seven GSI's were exposed to an integrated teacher training program (TTP), consisting of student evaluations and feedback to the instructors, videotaped classroom observations, and instructional seminars, which had been developed for training and assisting graduate student teachers. Tests on pre-test scores and student attributes indicated that the control (fall 1971) and experimental (winter 1972) groups of students were similar.

Mean post-*TUCE* scores and change-in-*TUCE* scores were significantly higher for

the winter quarter students than for the fall quarter students. Multiple linear regression analysis holding constant pre-*TUCE*, *ACT* score, grade point average, age, sex, and instructor's evaluation score confirmed the superior performance of students when the TTP was instituted. The data also revealed that the GSI training system had a significant favorable influence on instructors' ratings (on the Purdue Rating Scale).

The main problem with the Lewis-Orvis analysis is the possibility that the increased experience of the GSI's from fall to winter quarters is what actually caused the improvement in their teaching performance rather than the TTP. Otherwise the controls between the fall and winter quarter were excellent.

The evidence from the Minnesota experience, as well as the relatively better performance of GSI's at those schools that offer more advice on teaching, seems to suggest that at least a moderate amount of training may be effective.

The costs of GSI training programs appear to be quite modest. The Minnesota program, as well as other programs developed at Wisconsin, Harvard, Purdue, and a dozen or so other universities, requires relatively little graduate student time. To further reduce the costs of teacher training of GSI's, the JCEE, with the financial assistance of the Sloan Foundation, has developed a *Resource Manual for Teacher Training Programs in Economics* [137, Saunders, Welsh, and Hansen, 1978], which provides the materials necessary for conducting a TTP workshop.

V. *Lasting Effects of Economics Education*

The value of liberal education does not consist solely, or even mainly, of the knowledge retained. But the case for special emphasis on economics rests on its high marginal social product in a mixed economy, the functioning of which de-

pends on government policies strongly affected by public opinion.⁴⁰ The extent to which economics training has lasting effects is, therefore, an important question to investigate.

That a college course in elementary economics has *any* lasting impact on the ability of citizens to deal intelligently with policy issues was challenged by Stigler, who advanced an interesting hypothesis: "Select an adequate sample of seniors (I would prefer men five years out of college), equally divided between those who have never had a course in economics and those who have had a conventional one-year course. Give them an examination on current economic problems, not on textbook questions. I predict they will not differ in their performance"⁴¹ [159, 1963, p. 657]. The kind of examination Stigler had in mind was not a current events quiz but an analytic exercise.

Two major studies and some minor ones⁴² have shed light on the lasting ef-

fects of college economics courses. In the case of Bach and Saunders, the findings were part of a broader study [13, 1965]. In a multiple regression analysis of data from a 25-item version of the *TEU* taken by 3,966 high school teachers of social studies, Bach and Saunders found that having taken one or two economics courses in college did not add significantly to the teachers' score [13, 1965, p. 349]. Having taken three or four college courses added an amount that was statistically significant but quantitatively small relative to other variables. Five or more courses added a quantitatively significant amount to the score—but not nearly as much as having watched the television course, "The American Economy," three or more times a week during the preceding year. The test instrument used was not appropriate for investigating Stigler's hypothesis.

The one major study devoted mainly to lasting effects was carried out by Saunders [129, 1971; 132, 1973]. He used multiple regression analysis to compare three pairs of students or alumni. In each pair, one group had had an introductory college course in economics whereas the other had not. One pair included a set of sophomores who had just completed the introductory course and a set who had not; the second of seniors, one group of which had had the course two years earlier; the third of alumni five years after graduation. The measure of output was a hybrid version of the *Test of Understanding in College Economics* consisting of 33 questions cho-

⁴⁰ In Section II.C above, McKenzie was quoted as saying that people do not have enough incentive to maintain the human capital acquired through economic education and in any event will vote their own private interests [103, 1977, p. 10]. But an education should develop intellectual interests and at least lead people to a rational understanding of what their private interests are.

⁴¹ This hypothesis was the outgrowth of Stigler's criticism of the elementary course quoted in the introductory section of this survey.

⁴² A number of studies have been made of the lasting effects of high school economics (Bach and Saunders [14, 1966]; Moyer and Paden [111, 1968]; Saunders [128, 1970]; Weidenaar and Joe A. Dodson, Jr. [170, 1972]; C. D. Harbury and R. Szreter [63, 1968]; Attiyeh and Lumsden [9, 1971; 10, 1972]; Palmer *et al.* [120, 1979]). Stuart Wells characterized the results as "inconclusive" [172, 1974, p. 9]. Saunders and Bach in a study of 96 seniors who had taken a one-semester required course in economics as sophomores at Carnegie-Mellon University found that for those who took no more economics, the average score on the *TEU* declined a little over 10 percent *ceteris paribus* (about half the value added by the introductory course) but was still markedly higher than before taking any economics [135, 1970]. As Bach and Saunders noted, the *TEU* is extremely simple relative to the content of the Carnegie-Mellon course.

In a forthcoming paper, Andrew I. Kohen and Paul

H. Kipps estimate the decay rate of microeconomic knowledge acquired in an introductory course at James Madison University as 20 percent per year [85, 1979]. In another forthcoming paper, Eleanor D. Craig, James B. O'Neill, and Douglas W. Elfner of the University of Delaware find that class size did not affect retention [38, 1978]. (This is contrary to an earlier finding by Craig and O'Neill [37, 1978]). They also report that "the juniors who were freshmen during the course, outperformed the seniors (who had been sophomores)" [38, 1978, p. 2].

sen from the 132 in the *TUCE* to avoid technical analysis and specialized terminology. Twenty-three schools participated with a total of usable responses of 1,220 for sophomores, 955 for seniors, and 1,257 for alumni. Each respondent answered a detailed questionnaire, which provided information about variables hypothesized to be associated with differences in performance on the hybrid *TUCE*, interest in economics, and reading habits.

Saunders found that introductory economics courses did have a lasting impact on test performance. It diminished over time. Other things remaining the same, sophomores with an introductory course in economics scored 6.18 points higher on the hybrid *TUCE* than comparable sophomores without such a course; seniors who had had such a course two years previously scored 4.76 points higher than the corresponding control group; and alumni five years out of college scored 3.24 points higher if they had had an introductory course. Other things constant, each letter grade in introductory economics was significantly associated with a difference in total *TUCE* score of 2.00 for sophomores, 1.69 for seniors, and 1.09 for alumni. Each course taken beyond the introductory level was associated with a difference in test performance of .53 points in the senior sample and .48 in the alumni sample. A difference of one letter grade in each such course was associated with test score differences of .18 and .15 for seniors and alumni respectively. Among variables consistently and significantly associated with test scores in all three samples were the "intellectualism" of a school's student body, general interest of the person in economics as a subject, and reading the economics or business section of a weekly news magazine. Introductory economics courses and course grades did not appear to have a lasting impact on reported general interest in economics as a subject.

Two limitations of Saunders's study should be noted. Since he used a cross section of three sample groups at a single time rather than a longitudinal survey, the results may be suspect if the content or quality of the courses changed from the alumni group to the sophomores sampled. The use of questionnaires mailed to alumni has the usual response bias.

Despite the limitations, Saunders's study is one of the most important in the field of economics education. Though it did not test Stigler's hypothesis directly, it clearly implies that introductory college courses do result in a lasting increase in economics understanding. The difference between his results and the earlier study of Bach and Saunders may be attributed to the fact that one-third of those in the earlier study took their economics in teachers colleges, where the courses may have been inferior [13, Bach and Saunders, 1965, pp. 351-52]. It may also be associated with the different test instrument used, though the results of Saunders and Bach [135, 1970] indicate otherwise. As Bach and Saunders said of their results, "These findings emphasize again the well-known psychological principle that 'learning' unsupported by motivation and reinforcement . . . has a very short half-life" [13, 1965, p. 354].

VI. Summary and Conclusions

A. What Have We Learned?

The research on teaching college economics is voluminous. A substantial amount has been of good quality, some of it of high quality. So what have we learned?

- Different students learn economics in different ways. The best teaching strategy provides alternative learning methods directed toward the different needs of different students.
- Of the new teaching methods, the most effective seems to be computer-study-

management programs. In this method, the students take frequent tests and are given different assignments depending on the test results.

- Programmed learning is efficient in the sense of bringing students to a given level of competence in less time, but generally students do not like it.

- Students like self-paced instruction, and it increases learning in some circumstances.

- Graduate students generally are just as good teachers as regular faculty even though, other things equal, experience results in better teaching.

- Graduate students who have had teacher training are better instructors than those who have not.

- A one-year course in elementary economics has lasting effects in the form of greater economic competency.

- Computerized games may be fun but they do not seem to be worth the cost.

*B. How Can Research Methodology Be Improved?*⁴³

Modeling. Economics education is groping for a formal model, or set of models, of the instructional process. In the past much research in the field consisted of a college instructor regressing final examination scores for students on an ad hoc set of easily obtained student characteristic variables and a variable indicating whether the student participated in a certain pedagogical experiment. Little thought was given to model specification, and functional form was determined arbitrarily.

More recently, McKenzie and Staaf [104, 1974] and Daniel Graham and Kelley [55, 1974] have made progress toward exploiting the rich theoretical models of microeconomics, which explicitly consider behavioral reactions of students and

faculty to changes in prices and budget constraints. The application of such models promises to bear fruit, as the empirical literature on student and faculty behavior has already indicated. However, a substantial gap remains between theorizing and empirical research. Progress can be made by synthesizing and integrating the empirical work with the insights that are available from comprehensive theoretical models based on maximizing behavior.

Many of the alleged independent variables in economics education production function studies are endogenous. This has been recognized recently (Allison [2, 1976; 3, 1977]; Soper [153, 1976]; Craig Swan [160, 1978]; Becker and Salemi [18, 1977]), but most research conclusions continue to rest on single equation models. There are studies that use grades to predict course evaluation ratings, course evaluation ratings to predict grades, cognitive performance to predict course evaluation ratings, course evaluation ratings to predict cognitive performance, and grades to predict cognitive performance; and while there do not seem to be any explicit studies, everyone presumes (hopes) that cognitive performance affects grades. The need to develop a simultaneous equations model of this process is obvious.

Most of the evaluations of teaching innovations have been specified so that the treatment group enters the model as a shift parameter. This ignores the possibility that the effect of the alternative technique may be to improve the marginal product of some other factor that is already included in the equation. For example, by using separate equations, Siegfried and Strand [144, 1977] found that self-paced instruction was more effective for females than for males. Utilizing Blinder's method [19, 1973], Chizmar, Hiebert, and McCarney [31, 1977] found that users of computer-assisted instruction learned more economics in spite of CAI (rather than because of it).

⁴³ An excellent survey of conceptual and empirical issues in estimating educational production functions is Eric Hanushek [62, 1979].

One of the better modeling efforts is the Harvard project [3, Allison, 1977]. It constructs and estimates a three-equation model that adapts to the non-linearities and simultaneity inherent in the educational process. The model includes an equation describing student decisions about the allocation of time and effort; a production function equation that relates student effort, ability, and pedagogy to achievement; and a "profit function," relating student effort and achievement to student enjoyment of the course. Estimates from a constant elasticity of substitution form of the production function reveal "elasticities" of achievement with respect to student ability, pedagogic inputs, and effort of roughly .89, .40, and .25.

Sample Design. Research in economics education would be more credible if evaluation of innovative teaching technologies were conducted by individuals other than those who devise the new methods.

Too many studies in economics education have been one-school, one-time experiments. There are examples of large, broad-based data sets (Attiyeh, Bach, and Lumsden [6, 1969]; Attiyeh and Lumsden [9, 1971]; Saunders [132, 1973]), but they are rare. The single institution studies do not provide sufficient observations and sufficient variation in many variables required to disentangle the complicated relationships inherent in the educational process.

The main deficiency of the evaluations of innovative learning technologies, however, may involve the assignment of students to experimental and control groups. One rationale for the identification and evaluation of alternative pedagogical techniques is that efficient student-learning processes vary across individuals and the opportunity to select among alternative methods of learning permits students to choose the technique best suited to each individual, thus facilitating efficiency

in student learning. This suggests that the ultimate objective is not to entirely replace the conventional lecture format with alternative methods, but rather to offer options to (presumably well-informed) students.⁴⁴ If this is true, then the gain in learning will accrue to those students who would elect the experimental option and learn more in their courses than they would have learned in the conventional format had the experimental option been unavailable. Therefore, the experimental group should consist of students who voluntarily elect the experimental method and the *control group should consist of students who would voluntarily elect the experimental technology but are actually in the conventional course.*

Neither of the two methods typically used to identify control and experimental students—random assignment or voluntary self-selection—satisfies these criteria. With *randomly assigned* experimental and control students, some individuals are assigned to the experimental class who would have elected the conventional class with unconstrained choice. This implies that they believed *a priori* that the conventional format was more conducive to their learning of economics. Such people would not be found in the experimental course if it were offered as a regular option, and their inclusion in the experimental group confounds the empirical test. A similar argument can be made with respect to students who would elect the alternative technology if confronted with choice, but who are assigned (randomly) to the conventional format in the experiment. The performance of both groups may be worse than it would be if all students could select the personally most

⁴⁴ On the other hand, economies of scale relative to demand at some colleges may dictate the choice of only one pedagogical technique for economics instruction, in which case the usual random assignment of students to experimental and control groups is a satisfactory procedure.

suitable technology. Unfortunately, biases may not balance, and the empirical results may consequently provide misleading signals as to how useful the alternative technology would be *vis-à-vis* the conventional format for those students who would elect the new method under free choice.⁴⁵

If (informed) students are allowed to voluntarily choose whether to enroll in the experimental or control classes, we can infer that those choosing the alternative technology believe *a priori* that it would be helpful to them. So the experimental group includes the correct individuals. However, the control group then consists of students who have revealed that they would not elect the alternative method even if it were offered to them (as it was). Thus the control group does not provide any guidance as to how those who elect the new method would have performed in conventional classes if the alternative had not been offered.

In experiments to evaluate alternative pedagogical techniques it is necessary to identify students who would choose the experimental technology, and then divide them into an experimental and control group.⁴⁶

Measuring Outputs and Inputs. Many of the variables used to measure either inputs or outputs are confined to a small number of values. For example, usually there are only five possible course grades. Course evaluations commonly use a scale of 1 to 5. Often the underlying basis for such variables is an ordinal ranking, in which case (for independent variables) a

series of binary variables would be superior to arbitrarily imposing a cardinal relationship on the ranking (e.g., assuming that an A is worth twice a C).

A more serious problem, however, haunts the measurement of output and input. Many variables have upper and lower limits (e.g., ceiling and floor effects of "gap-closing" measures of cognitive achievement), in which case the error distribution is truncated, causing heteroskedasticity and biased parameter estimates if ordinary least squares regression is used [70, Thomas Johnson, 1979]. The practical importance of this limitation has recently been demonstrated by Lee Spector and Michael Mazzeo [157, 1978] who, in a study of self-paced instruction, found that the effect of a PSI introductory course on student grades in a later course was statistically significant and positive using ordinary least squares. When they adopted the more appropriate probit estimation technique, the null hypothesis (no effect of PSI) could no longer be rejected. There are now available several economical techniques for handling limited-value dependent variables [70, Johnson, 1979].

The major difficulty, however, with the measurement of outputs is the failure of most economics education research to recognize that students may elect to use their efficiency gain from a more efficient method of teaching economics to "purchase" additional knowledge in some other discipline (or perhaps "purchase" leisure). Such gains are overlooked when conventional measurement techniques are used. The empirical research on study time comes closest to recognizing this problem.

Multicollinearity. The ad hoc nature of most empirical investigations of the production function for economics education has blurred the distinction between conceptualizing a model and measuring the variables in it. Consequently multiple measures of single concepts often appear

⁴⁵ A rigorous demonstration of the sample bias in most economics education research is contained in Siegfried and George Sweeney [145, 1979].

⁴⁶ This experimental design runs the risk of bias from a "negative Hawthorne effect." There is danger that students assigned to the control group, knowing and preferring the experimental pedagogy, might perform below normal because they are dissatisfied at being denied the opportunity to learn in the experimental course.

in one equation, causing multicollinearity and substantive errors in interpretation.⁴⁷ Controlling for all of the obvious factors that learning theory predicts to affect output is important for reducing bias in the estimated parameters. However, a decision must be made when there are duplicate measures available for one factor.

If and when theory provides little or no guidance as to which of several alternative measures is superior and more than one is included in the model, then the group of measures should be tested jointly with an *F*-test. It is likely that this approach would explain the apparent paradox of so many standard variables being insignificant in empirical studies of economics education. A joint *F*-test on a group of variables is also appropriate when a change in one variable necessarily causes a change in another variable, since the impact of changing the one variable alone is then irrelevant. For example, if an achievement equation contains a binary variable for an experimental teaching method and an interaction term that is the product of the experimental binary variable and, say, SAT scores, it makes no sense to consider the separate effect of the first variable alone, since changing the teaching method would necessarily change the value of the interaction term also.

Multicollinearity has clearly been recognized in economics education research (Soper [150, 1973; 153, 1976]; Becker [16, 1976]; Highsmith [67, 1976]; Swan [160, 1978]). The continuing debate centers on

what to do about it.⁴⁸ In addition to eliminating extraneous measures of single factors and jointly testing multiple measures of a single factor for statistical significance, it may be possible to incorporate new information into the model to circumvent the problem. For example, if in a time-series regression of Y on X_1 and X_2 , X_1 and X_2 are highly collinear, one might turn to cross-sectional data where, luckily, X_1 is constant across observations. The relationship between X_2 and Y can then be estimated from the cross-sectional data and the time-series analysis used to estimate the coefficient of X_1 while constraining the coefficient of X_2 to the value determined from the cross-sectional analysis. Another new technique for dealing with multicollinearity is "ridge regression." Its advantages, however, come at considerable cost [70, Johnson, 1979].⁴⁹

Our criticisms of research methodology certainly do not apply without exception. Many studies have handled these problems well. For example, Attiyeh, Bach, and Lumsden employed a sample of 4,121 students at 48 different colleges and universities to estimate a production function for economics education [6, 1969]. Allison devised and estimated a simultaneous equations model of the education production function for introductory economics at Harvard using over 2,400 student observations [3, 1977]. Her study included a model of student time allocation based on utility maximization and allowed for non-linearities (*i.e.*, diminishing marginal returns to certain factors) and

⁴⁷ Describing production function studies of education in general, Hanushek argues that "... multicollinearity does not appear to be the villain it has been made out to be, although it may partially explain some of the apparent inconsistencies in existing research. ... The usual terminology for regression analysis is misleading here: Right hand variables are often called independent variables, but this does not imply that they cannot be correlated. In fact, multiple regression analysis is used *because* there are correlations among the 'independent' variables" [62, 1979, pp. 351-88].

⁴⁸ Some of the earlier studies of economics education production functions occasionally used stepwise regression to choose among a plethora of variables. The level of sophistication of research methods has now taken us beyond such thoroughly discredited techniques for hypothesis testing.

⁴⁹ Ridge regression changes some of the (diagonal) values in the regression matrix. This reduces the standard errors of the estimated coefficients, providing narrower confidence intervals. However, it also introduces unknown biases into the point estimates of the parameters.

differential effects on different types of students. Becker and Salemi evaluated an audiovisual tutorial package using data from six colleges [18, 1977]. They explicitly considered a nonlinear theoretical model of learning; modeled the ceiling effect of the gap-closing cognitive achievement measure; corrected for simultaneous equations bias; controlled for variable inputs (*i.e.*, student study time); specified interaction terms, which permitted the identification of beneficiaries of the tutorial package; and analyzed the marginal costs of the program and attempted to weigh them against the marginal benefits.

The field of economics education is a teenager, experiencing the growing pains of adolescence. The rapid improvement in research methodology being applied to problems of teaching in economics should not be surprising in view of this youth. Indeed, what is remarkable is the progress that has been made in such a relatively short span.

C. Where Do We Go From Here?

There are two major thrusts that should be undertaken to complement the improvements in research methods discussed above. First, many of the good one-school studies need to be replicated elsewhere. The evidence from Attiyeh, Bach, and Lumsden [6, 1969], that different types of schools matter, and from Chizmar *et al.* [31, 1977], that the effect of teaching methods on economics learning may depend on the level of human capital available, suggests the importance of replication. In addition, there is danger that the innovator invested substantially more time in teaching an experimental course, in which case advantages attributed to a new method actually may be returns to faculty effort. This danger is reduced if the innovation is implemented and evaluated by someone other than the initiator. Finally, replication will add to the sample size and improve confidence in the objec-

tivity of assessments of teaching methods. The most effective means of gaining large sample sizes and sufficient variation in explanatory factors to permit sophisticated modeling while minimizing the danger of variation in the implementation of teaching methods is large-scale research projects similar to those of Attiyeh, Bach and Lumsden [6, 1969], Saunders [129, 1971; 132, 1973] or Attiyeh and Lumsden [9, 1971].

Second, most of the research in economics education has been concerned with the college principles of economics course, and rightly so. However, there are other important vehicles of economics education, and perhaps the early marginal products from an assessment of them exceed the diminished marginal product of yet another twist on programmed instruction, televised lectures, or computer games. For example, most of the economics education in the United States comes through the popular press. Very little assessment has been made of advertising campaigns by corporations, trade association education (propaganda?) efforts, or press coverage of economic events. What is the impact of the PBS series "Economically Speaking" or "The Age of Uncertainty"? What is the quality of economics reasoning among the business and economics staffs of local newspapers?

Besides the principles course there is the economics major. While there have been a few studies of upper division courses (economic statistics, intermediate theory, money and banking), for the most part these courses have been neglected by research. This may be due to the absence of clearly defined goals for the economics major, of which these courses are usually a part. Since undergraduate majors in economics pursue diverse careers—law school, business school, graduate school in economics (and other disciplines), business employment, government employment, entrepreneurship—it

is not obvious how the undergraduate major curriculum should be structured. Indeed, there is substantial disagreement on whether it should be career oriented at all.

Data on undergraduate instruction in economics at U.S. colleges and universities could be useful in planning and administering Ph.D. programs, since undergraduate majors are an important component in the derived demand for Ph.D.'s [138, Scott, 1979]. Before research proceeds to courses beyond elementary economics, we need to know how many people annually enroll in which courses and for what purposes.³⁰ The debate on course objectives, while unsettled for the principles course, may be highly controversial for the set of courses that constitute a major.

There are many other important areas for research on teaching college economics. For example, further refinement of the input coefficients in the production function will be useful as colleges and universities begin belt-tightening in response to the diminishing college age population in the next decade. Research findings may help college administrators to decide between reducing budgets by increasing class size or by hiring younger, and less experienced, instructors.

In spite of rapid growth in alternative pedagogies, the blackboard and textbook remain the staple inputs into the process. But we know embarrassingly little about how differences among textbooks affect the learning and attitudes of students. We have done little detailed analysis of the generalizable attributes of classroom teaching behavior. The JCEE's Teacher Training Program contains many helpful

hints to improve lectures, but to date there has been little, if any, systematic research to verify their effectiveness.

It appears that the standard production function studies of college teaching have been unable to explain much of the variation in measured outputs. This may be because the inputs or outputs have been measured inadequately, or because the model was specified poorly. An alternative explanation is that the important determinants of variations in student learning are at the micro level—individual students have such different learning processes that generalization is almost futile. From the more general research on education, Davis Armor *et al.* [5, 1976] explain the apparent ineffectiveness of alternative pedagogies by the substantial variation in implementation of them at the classroom level. Therefore, we need to learn more about the process by which instructors adopt teaching methods and tailor them to individual students.

D. Conclusion

A cumulative literature on economics education has now developed. As in other subfields of economics, those who would publish must search the literature for previous findings and build on them. They must also bring to bear the tools of economic theory and econometrics. The quality of the research done so far varies widely, but dramatic improvement has occurred in recent years; much more of the current research is first rate than was the case 15 years ago.

An economist can now make a decent living by specializing in economics education. True (perhaps unfortunately for undergraduate students), it is not a prestige subject in high demand. A young economist wanting to win the Clark medal needs to earn his reputation elsewhere. But the field is rapidly becoming respectable, and the research findings can be useful to college economics teaching.

³⁰ A count of professors teaching various courses by a company that sells mailing lists to publishers listed 6,130 for principles of economics, 1,671 for intermediate micro theory, 1,408 for money and banking, 1,404 for intermediate macro theory, and 859 for labor economics as the five most popular course areas in 1978-79 [33, College Marketing Group, n.d.].

REFERENCES

1. ALLISON, ELISABETH. "Self-Paced Instruction: A Review," *J. Econ. Ed.*, Fall 1975, 7(1), pp. 5-12.
2. ———. "Three Years of Self-Paced Teaching in Introductory Economics at Harvard," *Amer. Econ. Rev.*, May 1976, 66(2), pp. 222-28.
3. ———. "Educational Production Function for an Introductory Economics Course." Discussion Paper Number 545, April. Harvard University Institute of Economic Research, 1977.
4. AMERICAN ECONOMIC ASSOCIATION. "Classification of Members by Fields of Specialization," *Amer. Econ. Rev.*, December 1978, 68(6), pp. 426-31.
5. ARMOR, DAVIS J., ET AL. *Analysis of the school preferred reading program in selected Los Angeles minority schools*. R-20007-LAUSD. Santa Monica: Rand, 1976.
6. ATTIYEH, RICHARD; BACH, GEORGE L. AND LUMSDEN, KEITH G. "The Efficiency of Programmed Learning in Teaching Economics: The Results of a Nationwide Experiment," *Amer. Econ. Rev.*, May 1969, 59(2), pp. 217-23.
7. ———; BACH, GEORGE L. AND LUMSDEN, KEITH G. "Factors Affecting Student Learning of Elementary Economics," in *Recent research in economic education*. Edited by Keith G. Lumsden. Englewood Cliffs, N.J.: Prentice-Hall, 1970, pp. 67-79.
8. ——— AND LUMSDEN, KEITH G. "The Effectiveness of 'Programmed Learning' in Economics," *Amer. Econ. Rev.*, May 1965, 55(2), pp. 549-55.
9. ——— AND LUMSDEN, KEITH G. "University Students' Initial Understanding of Economics: The Contribution of the 'A' Level Economics Course and of Other Factors," *Economica*, Feb. 1971, 38(149), pp. 81-97.
10. ——— AND LUMSDEN, KEITH G. "Some Modern Myths in Teaching Economics: The U.K. Experience," *Amer. Econ. Rev.*, May 1972, 62(2), pp. 429-33.
11. BACH, GEORGE L. "Student Learning in Basic Economics: An Evaluated Experimental Course," in *New developments in the teaching of economics*. Edited by KEITH G. LUMSDEN. Englewood Cliffs, N.J.: Prentice Hall, 1967, pp. 74-91.
12. ———. "What Should a Principles Course in Economics Be?" in *Goals and objectives of the introductory college-level course in economics*. Edited by ALLEN LARSEN AND ANDREW T. NAPPI. Ninth District Economic Information Series. Minneapolis: The Federal Reserve Bank of Minneapolis, 1976, pp. 15-18.
13. ——— AND SAUNDERS, PHILLIP. "Economic Education: Aspirations and Achievements," *Amer. Econ. Rev.*, June 1965, 55(3), pp. 329-56.
14. ——— AND SAUNDERS, PHILLIP. "Lasting Effects of Economics Courses at Different Types of Institutions," *Amer. Econ. Rev.*, June 1966, 56(3), pp. 505-11.
15. BATTALIO, RAYMOND C.; HULETT, JOE R. AND KAGEL, JOHN H. "Comment on J. J. Siegfried's 'The Publishing of Economic Papers and Its Impact on Graduate Faculty Ratings, 1960-1969,'" *J. Econ. Lit.*, March 1973, 11(1), pp. 68-70.
16. BECKER, WILLIAM E., JR. "Programmed Instruction in Large-Lecture Courses: A Technical Comment," *J. Econ. Ed.*, Fall 1976, 8(1), pp. 38-40.
17. ———. "Professorial Behavior Given a Stochastic Reward Struc-

Siegfried and Fels: Teaching College Economics

- ture," *Amer. Econ. Rev.*, Dec. 1979, 69(5), forthcoming.
18. ——— AND SALEMI, MICHAEL K. "The Learning and Cost Effectiveness of AVT Supplemented Instruction: Specification of Learning Models," *J. Econ. Ed.*, Spring 1977, 8(2), pp. 77-92.
19. BLINDER, ALAN S. "Wage Discrimination: Reduced Form and Structural Estimates," *J. Human Res.*, Fall 1973, 8(4), pp. 436-55.
20. BONELLO, FRANK J.; DAVISSON, WILLIAM I. AND SWARTZ, THOMAS R. "Educational Innovations May Be Statistically Significant If Properly Measured," paper presented at the Allied Social Science Associations Annual Meeting. Chicago, August 1978.
21. BOOMS, BERNARD H. AND KALTREIDER, D. LYNNE. "Computer-Aided Instruction for Large Elementary Courses," *Amer. Econ. Rev.*, May 1974, 64(2), pp. 408-13.
22. BOULDING, KENNETH AND BOULDING, ELISE. "Introducing Freshmen to the Social System," *Amer. Econ. Rev.*, May 1974, 64(2), pp. 414-19.
23. BOWLES, SAMUEL. "Towards an Educational Production Function," in *Education, income, and human capital*. Edited by W. LEE HANSEN. New York: National Bureau of Economic Research; distributed by Columbia University Press, 1970, pp. 11-61.
24. BRONFENBRENNER, MARTIN. "Radical Economics in America: A 1970 Survey," *J. Econ. Lit.*, Sept. 1970, 8(3), pp. 747-66.
25. ———. "The Economics Major—What Is He?" *Atlantic Econ. J.*, Dec. 1978, 6(4), pp. 22-25.
26. BUCHELE, ROBERT AND LAZONICK, WILLIAM. "Economics as a Social Science: Introducing the Capitalist Economy," *Rev. Rad. Pol. Econ.*, Winter 1975, 6(4), pp. 20-40.
27. BUCKLES, STEPHEN G. AND MCMAHON, MARSHALL E. "Further Evidence on the Value of Lectures in Elementary Economics," *J. Econ. Ed.*, Spring 1971, 2(2), pp. 138-41.
28. ——— AND WELSH, ARTHUR L. "The Use of Validated Tests in Teaching and Research," in WELSH, ARTHUR L., ed. [173, 1972], pp. 24-32.
29. CAPOZZA, DENNIS R. "Student Evaluations, Grades and Learning in Economics (912)," *Western Econ. J.*, 11(1), March 1973, p. 127.
30. CASPER, CHERYL A. "An Empirical Study of the Effect of Behavioral Objectives in Principles of Economics," *J. Econ. Ed.*, Spring 1977, 8(2), pp. 127-29.
31. CHIZMAR, JOHN F.; HIEBERT, L. DEAN AND MCCARNEY, BERNARD J. "Assessing the Impact of an Instructional Innovation on Achievement Differentials: The Case of Computer-Assisted Instruction," *J. Econ. Ed.*, Fall 1977, 9(1), pp. 42-46.
32. COHN, ELCHANAN. *The economics of education*. Revised Edition. Cambridge, Massachusetts: Ballinger Publishing Company, [1972] 1979.
33. COLLEGE MARKETING GROUP, INC. *CMG Mailing Lists 1978-1979*. Winchester, Mass., no date.
34. COMMITTEE FOR ECONOMIC DEVELOPMENT, NATIONAL TASK FORCE ON ECONOMIC EDUCATION. *Economic education in the schools: A report*. New York, CED, 1961.
35. COSTIN, FRANK; GREENOUGH, WILLIAM T. AND MENGES, ROBERT J. "Student Ratings of College Teaching: Reliability, Validity, and Usefulness," *J. Econ. Ed.*, Fall 1973, 5(1), pp. 51-53.
36. COX, STEVEN R. "Computer-Assisted Instruction and Student Performance in Macroeconomic Princi-

- ples," *J. Econ. Ed.*, Fall 1974, 6(1), pp. 29-37.
37. CRAIG, ELEANOR D. AND O'NEILL, JAMES B. "Economic Education: Where Are We Going?" paper presented to the Eastern Economic Association, Washington, D.C., April 1978.
38. ———; O'NEILL, JAMES B. AND ELFNER, DOUGLAS W. "Large Class Retention: The Effects of Method in Macroeconomics." Unpublished manuscript, University of Delaware, September 1978.
39. DARTON, DONALD C. "Programmed Learning-Policy Analysis: An Experiment in Teaching Principles of Economics," *J. Econ. Ed.*, Fall 1971, 3(1), pp. 32-35.
40. DAVISSON, WILLIAM I. AND BONELLO, FRANK J. *Computer-assisted instruction in economics education: A case study*. Notre Dame, Indiana: University of Notre Dame Press, 1976.
41. EDWARDS, RICHARD C. AND MAC- EWAN, ARTHUR. "A Radical Approach to Economics: Basis for a New Curriculum," *Amer. Econ. Rev.*, May 1970, 60(2), pp. 352-63.
42. ELLIOTT, RALPH D.; IRELAND, M. EDWIN AND CANNON, TERESA S. "An Analysis of the Marginal Products of the One- and Two-Semester Introductory Economics Courses," *J. Econ. Ed.*, Fall 1978, 10(1), pp. 18-25.
43. EMERY, E. DAVID AND ENGER, THOMAS P. "Computer Gaming and Learning in an Introductory Economics Course," *J. Econ. Ed.*, Spring 1972, 3(2), pp. 75-85.
44. ——— AND SCHOENE, JEAN. "Computer Gaming and Learning: A Note on Retention." Unpublished manuscript, St. Olaf College, circa 1974.
45. EVERETT, MICHAEL D. "Student Evaluations of Teaching and the Cognitive Level of Economics Courses," *J. Econ. Ed.*, Spring 1977, 8(2), pp. 100-103.
46. FELS, RENDIGS. "On Teaching Elementary Economics," *Amer. Econ. Rev.*, Dec. 1955, 45(5), pp. 919-32.
47. ———. "A New Test of Understanding in College Economics," *Amer. Econ. Rev.*, May 1967, 57(2), pp. 660-66.
48. ———. "Multiple Choice Questions in Elementary Economics," in *Recent research in economics education*. Edited by KEITH G. LUMSDEN. Englewood Cliffs, N.J.: Prentice-Hall, 1970, pp. 27-43.
49. ———. *The Vanderbilt-JCEE experimental course in elementary economics*. (*J. Econ. Ed.*, Winter 1974, Special Issue Number 2.) New York: Joint Council on Economic Education, 1974.
50. ———. "Who's Been Killing the Free Enterprise System?" *Vanderbilt Alumnus*, Spring 1978, pp. 8-9.
51. ——— AND STARLEAF, DENNIS R. "Controlled Experiments in Teaching Techniques," *Southern Econ. J.*, July 1963, 30(1), pp. 68-73.
52. FREEMAN, RICHARD B. *The market for college-trained manpower: A study in the economics of career choice*. Cambridge, Mass.: Harvard University Press, 1971.
53. FREIDEN, ALAN AND STAAB, ROBERT J. "Scholastic Choice: An Economic Model of Student Behavior," *J. Human Res.*, Summer 1973, 8(3), pp. 396-404.
54. FUSFELD, DANIEL R. AND JUMP, GREGORY. "An Experiment with Programmed Instruction in Economics," *Southern Econ. J.*, Jan. 1966, 32(3), pp. 353-56.
55. GRAHAM, DANIEL A. AND KELLEY, ALLEN C. "Towards a Behavioral Model of the Educational Process."

Siegfried and Fels: Teaching College Economics

- Unpublished manuscript, Duke University, October 1974.
56. GURLEY, JOHN G. "Some Comments on the Principles Course," *Amer. Econ. Rev.*, May 1975, 65(2), pp. 431-33.
57. HANSEN, W. LEE. "New Approaches to Teaching the Principles Course," *Amer. Econ. Rev.*, May 1975, 65(2), pp. 434-37.
58. ———. "The State of Economic Literacy," in WENTWORTH, DONALD R.; HANSEN, W. LEE AND HAWKE, SHARRYL H., eds. [175, 1977], pp. 61-79.
59. ———, et al. *Master curriculum guide for the nation's schools. Part I: A framework for teaching economics: Basic concepts*. New York: Joint Council on Economic Education, 1977.
60. ——— AND KELLEY, ALLEN C. "Political Economy of Course Evaluations," *J. Econ. Ed.*, Fall 1973, 5(1), pp. 10-21.
61. ———; KELLEY, ALLEN C. AND WEISBROD, BURTON A. "Economic Efficiency and the Distribution of Benefits From College Instruction," *Amer. Econ. Rev.*, May 1970, 60(2), pp. 364-69.
62. HANUSHEK, ERIC A. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *J. Human Res.*, Summer 1979, 14(3), pp. 351-88.
63. HARBURY, C. D. AND SZRETER, R. "The Influence upon University Performance of the Study of Economics at School," *J. Royal Statist. Soc., Series A (General)*, 1968, 131, Part 3, pp. 384-409.
64. HARTMAN, DAVID G. "What Do Economics Majors Learn?" *Amer. Econ. Rev.*, May 1978, 68(2), pp. 17-22.
65. ———. "What Do Economics Majors Learn?" *J. Econ. Ed.*, Spring 1978, 9(2), pp. 87-91.
66. HAVRILESKY, THOMAS. "A Test of the Effectiveness of Teaching Money and Banking by Programmed Instruction," *J. Econ. Ed.*, Spring 1971, 2(2), pp. 151-54.
67. HIGHSMITH, ROBERT. "Second Generation Research in Economic Education: A Comment," *J. Econ. Ed.*, Fall 1976, 8(1), pp. 48-51.
68. HUANG, CLIFF J. "Optimal Allocation of Faculty Time under Uncertainty in Production." Unpublished manuscript, Vanderbilt University, February 1, 1979.
69. JENKINS, EVANS. "Should Cincinnati Bell Charge Extra? A Case-Study Approach to Economics," in "Report on Teaching 3," *Change*, Jan. 1977, 9(1), pp. 8-10.
70. JOHNSON, THOMAS. "Research on Economic Education: How Well Is It Answering the Questions Asked?" *Amer. Econ. Rev.*, May 1979, 69(2), pp. 22-27.
71. JOINT COUNCIL ON ECONOMIC EDUCATION. *Test of Economic Understanding*. Chicago: Science Research Associates, 1963.
72. ———. *Test of Understanding in College Economics*. New York: Psychological Corp., 1968.
73. JOSEPH, MYRON L. "Game and Simulation Experiments," *J. Econ. Ed.*, Spring 1970, 1(2), pp. 91-96.
74. KARSTENSSON, LEWIS AND VEDDER, RICHARD K. "A Note on Attitude as a Factor in Learning Economics," *J. Econ. Ed.*, Spring 1974, 5(2), pp. 109-11.
75. KATZ, DAVID A. "Faculty Salaries, Promotion, and Productivity at a Large University," *Amer. Econ. Rev.*, June 1973, 63(3), pp. 469-77.
76. KAU, JAMES B. AND RUBIN, PAUL H. "Measurement Techniques, Grades and Ratings of Instructors," *J. Econ. Ed.*, Fall 1976, 8(1), pp. 59-62.
77. KELLER, FRED S. "Goodbye Teacher

- ..., " in *PSI: Personalized system of instruction: 41 germinal papers*. Edited by J. GILMOUR SHERMAN. Menlo Park, Calif.: W. A. Benjamin, 1974, pp. 3-12. (See also pp. 209-15 for two related articles also by Keller.)
78. KELLEY, ALLEN C. "An Experiment with TIPS: A Computer-Aided Instructional System for Undergraduate Education," *Amer. Econ. Rev.*, May 1968, 58(2), pp. 446-57.
 79. ———. "The Economics of Teaching: The Role of TIPS," in *Recent research in economics education*. Edited by KEITH G. LUMSDEN. Englewood Cliffs, N.J.: Prentice-Hall, 1970, pp. 44-66.
 80. ———. "TIPS and Technical Change in Classroom Instruction," *Amer. Econ. Rev.*, May 1972, 62(2), pp. 422-28.
 81. ———. "Uses and Abuses of Course Evaluations as Measures of Educational Output," *J. Econ. Ed.*, Fall 1972, 4(1), pp. 13-18.
 82. ———. "Individualizing Education through the Use of Technology in Higher Education," *J. Econ. Ed.*, Spring 1973, 4(2), pp. 77-89.
 83. ———. "The Student as a Utility Maximizer," *J. Econ. Ed.*, Spring 1975, 6(2), pp. 82-92.
 84. KIPPS, PAUL H. "The Use of Course Evaluation Scores to Influence Teaching and Research Activities," *J. Econ. Ed.*, Spring 1975, 6(2), pp. 93-98.
 85. KOHEN, ANDREW I. AND KIPPS, PAUL H. "Factors Determining Students' Retention of Economic Knowledge after Completing the Principles-of-Microeconomics Course," *J. Econ. Ed.*, Spring 1979, 10(2).
 86. LABINSKI, P. F. "The Effectiveness of Economics Instruction in Two-Year Colleges Revisited," *J. Econ. Ed.*, Spring 1978, 9(2), pp. 102-06.
 87. LAMPHEAR, CHARLES AND MCCONNELL, CAMPBELL R. "A Note on the Use of Graduate Teaching Assistants in the Principles Course," *J. Econ. Ed.*, Spring 1970, 1(2), pp. 139-42.
 88. LEFTWICH, RICHARD H. "Objectives of the College-Level Principles of Economics Course," in *Goals and objectives of the introductory college-level course in economics*. Edited by ALLEN LARSEN AND ANDREW T. NAPPI. Ninth District Economic Information Series. Minneapolis: The Federal Reserve Bank of Minneapolis, 1976, pp. 26-29.
 89. ——— AND SHARP, ANSEL M. *Syllabus for an 'issues approach' to teaching economics principles*. (*J. Econ. Ed.*, Winter 1974, Special Issue Number 1.) New York: Joint Council on Economic Education, 1974.
 90. LEVIN, HARRY M. *Differences in outcomes between large and small classes in Western Civilization and Economics*. Ph.D. dissertation, Rutgers University, UM Number 67-14, 428, 1967.
 91. LEWIS, DARRELL R. AND DAHL, TOR. "The Test of Understanding in College Economics and Its Construct Validity," *J. Econ. Ed.*, Spring 1971, 2(2), pp. 155-66.
 92. ——— AND DAHL, TOR. "Critical Thinking Skills in the Principles Course: An Experiment," in *Research papers in economic education*. WELSH, ARTHUR L., ed. [173, 1972], pp. 50-68.
 93. ——— AND DAHL, TOR. "Factors Influencing Performance in the Principles Course Revisited," in *Research papers in economic education*. WELSH, ARTHUR L., ed. [173, 1972], pp. 94-117.
 94. ——— AND ORVIS, CHARLES C. "A Training System for Graduate Student Instructors of Introductory Eco-

- nomics at the University of Minnesota," *J. Econ. Ed.*, Fall 1973, 5(1), pp. 38-46.
95. LUMSDEN, KEITH G. "The Effectiveness of Programmed Learning in Elementary Economics," *Amer. Econ. Rev.*, May 1967, 57(2), pp. 652-59.
96. ———. "The Information Content of Student Evaluation of Faculty and Courses," in *Efficiency in universities: The La Paz papers*. Edited by KEITH G. LUMSDEN. Amsterdam and New York: Elsevier Scientific, 1974, pp. 175-204.
97. ———. ATTIYEH, RICHARD E. AND SCOTT, ALEX. *Economics education in the U.K.* Heinemann Educational Books, forthcoming.
98. MANN, WILLIAM R. AND FUSFELD, DANIEL R. "Attitude Sophistication and Effective Teaching in Economics," *J. Econ. Ed.*, Spring 1970, 1(2), pp. 111-29.
99. MCCONNELL, CAMPBELL R. "An Experiment with Television in the Elementary Course," *Amer. Econ. Rev.*, May 1968, 58(2), pp. 469-82.
100. ——— AND FELTON, JOHN R. "A Controlled Evaluation of 'The American Economy,'" *Amer. Econ. Rev.*, June 1964, 54(3), pp. 403-07.
101. ——— AND LAMPHEAR, CHARLES. "The Effectiveness of Teaching Methods: Teaching Principles of Economics without Lectures," *J. Econ. Ed.*, Fall 1969, 1(1), pp. 20-32.
102. MCKENZIE, RICHARD B. "The Economic Effects of Grade Inflation on Instructor Evaluations: A Theoretical Approach," *J. Econ. Ed.*, Spring 1975, 6(2), pp. 99-106.
103. ———. "Where Is the Economics in Economic Education?" *J. Econ. Ed.*, Fall 1977, 9(1), pp. 5-13.
104. ——— AND STAAF, ROBERT J. *An economic theory of learning: Student sovereignty and academic freedom*. Blacksburg, Va.: University Publications, 1974.
105. MEEROPOL, MICHAEL. "A Radical Teaching a Straight Principles of Economics Course," *Rev. Rad. Pol. Econ.*, Winter 1975, 6(4), pp. 2-9.
106. MIRUS, ROLF. "Some Implications of Student Evaluation of Teachers," *J. Econ. Ed.*, Fall 1973, 5(1), pp. 35-46.
107. MORAWETZ, DAVID. *Helping students learn to identify and set up economic problems: An evaluated experiment*. Jerusalem: Hebrew University Economics Department Research Report Number 90, 1976.
108. ———. "The Use of Graduate Student Instructors in the Introductory Economics Course at the Hebrew University, 1967-1975" *J. Econ. Ed.*, Fall 1977, 9(1), pp. 47-50.
109. ———. "Helping Students to Learn to Identify and Set Up Economic Problems: An Evaluated Experiment," *J. Econ. Ed.*, Spring 1978, 9(2), pp. 92-95.
110. MORGAN, W. DOUGLAS AND VASCHÉ, JON DAVID. "An Educational Production Function Approach to Teaching Effectiveness and Evaluation," *J. Econ. Ed.*, Spring 1978, 9(2), pp. 123-26.
111. MOYER, M. EUGENE AND PADEN, DONALD W. "On the Efficiency of the High School Economics Course," *Amer. Econ. Rev.*, Sept. 1968, 58(4), pp. 870-76.
112. MURNANE, RICHARD J. *The impact of school resources on the learning of inner city children*. Cambridge, Mass.: Ballinger, 1975.
113. NEEDHAM, DOUGLAS. "Student Effort, Learning, and Course Evaluation," *J. Econ. Ed.*, Fall 1978, 10(1), pp. 35-43.
114. NELSON, DENNIS L. *The effect of specifically stated instructional objectives on the achievement of collegi-*

- ate undergraduate economics students. Ph.D. dissertation, University of Minnesota, 71-18, 789, 1970.
115. NICHOLS, ALAN AND SOPER, JOHN C. "Economic Man in the Classroom," *J. Polit. Econ.*, Sept.-Oct. 1972, 80(5), pp. 1069-73.
116. OATES, WALLACE E. AND QUANDT, RICHARD E. "The Effectiveness of Graduate Students as Teachers of Principles of Economics," *J. Econ. Ed.*, Spring 1970, 1(2), pp. 130-38.
117. O'CONNOR, WILLIAM C. "Economic Education: Comparative Experiences of Great Britain and the United States," *Economica: The Journal of the Economic Association*, Winter 1973-74, 10(3), pp. 171-74.
118. PADEN, DONALD W.; DALGAARD, BRUCE R. AND BARR, MICHAEL D. "A Decade of Computer-Assisted Instruction," *J. Econ. Ed.*, Fall 1977, 9(1), pp. 14-20.
119. ——— AND MOYER, M. EUGENE. "The Effectiveness of Teaching Methods: The Relative Effectiveness of Three Methods of Teaching Principles of Economics," *J. Econ. Ed.*, Fall 1969, 1(1), pp. 33-45.
120. PALMER, JOHN; CARLINER, GEOFFREY AND ROMER, THOMAS. "Does High School Economics Help?" *J. Econ. Ed.*, Spring 1979, 10(2).
121. PHILLIPS, JAMES A. "Instructional Objectives and Economics Understanding," *J. Econ. Ed.*, Spring 1972, 3(2), pp. 112-17.
122. RAMSETT, DAVID E.; JOHNSON, JERRY D. AND ADAMS, CURTIS. "Some Evidence on the Value of Instructors in Teaching Economic Principles," *J. Econ. Ed.*, Fall 1973, 5(1), pp. 57-62.
123. REMMERS, H. H. AND WEISBRODT, J. A. *Manual of instructions for the Purdue rating scale for instruction*. West Lafayette, Ind.: University Bookstore, 1965.
124. RODIN, MIRIAM AND RODIN, BURTON. "Student Evaluations of Teachers," *J. Econ. Ed.*, Fall 1973, 5(1), pp. 5-21.
125. ROKEACH, MILTON. *The nature of human values*. New York: Macmillan, Free Press, 1973.
126. ROTHMAN, MITCHELL P. AND SCOTT, JAMES H., JR. "Political Opinions and the TUCE," *J. Econ. Ed.*, Spring 1973, 4(2), pp. 116-24.
127. SAUNDERS, PHILLIP. "The Effectiveness of 'The American Economy' in Training Secondary School Teachers," *Amer. Econ. Rev.*, June 1964, 54(3), pp. 396-403.
128. ———. "Does High School Economics Have a Lasting Impact?" *J. Econ. Ed.*, Fall 1970, 2(1), pp. 39-55.
129. ———. "The Lasting Effects of Elementary Economics Courses: Some Preliminary Results," *Amer. Econ. Rev.*, May 1971, 61(2), pp. 242-48.
130. ———. "More on the Use of Graduate Student Instructors in the Introductory Economics Course," *J. Econ. Ed.*, Fall 1971 3(1), pp. 36-40.
131. ———. "Student Learning and Instructor Ratings: The Carnegie-Mellon Experience in Introductory Economics," in WELSH, ARTHUR L., ed. [173, 1972], pp. 2-11.
132. ———. "The Lasting Effectiveness of Introductory Economics Courses," final report, NSF Grant GY-7208, June 1973.
133. ———. *Experimental course development in introductory economics at Indiana University*. (*J. Econ. Ed.*, Fall 1975, Special Issue Number 4.) New York: Joint Council on Economic Education, 1975.
134. ———. "Instructional Objectives," in SAUNDERS, PHILLIP; WELSH, ARTHUR L. AND HANSEN, W. LEE, eds. [137, 1978], pp. 51-87.
135. ——— AND BACH, GEORGE L. "The Lasting Effects of an Introductory

Siegfried and Fels: Teaching College Economics

- Economics Course: An Exploratory Study," *J. Econ. Ed.*, Spring 1970, 1(2), pp. 143-49.
136. ——— AND WELSH, ARTHUR L. "The Hybrid TUCE: Origin, Data, and Limitations," *J. Econ. Ed.*, Fall 1975, 7(1), pp. 13-19.
137. ———; WELSH, ARTHUR L. AND HANSEN, W. LEE. *Resource manual for teacher training programs in economics*. New York: Joint Council on Economic Education, 1978.
138. SCOTT, CHARLES E. *The academic market for Ph.D. economists: An econometric analysis*. Ph.D. dissertation, Vanderbilt University, 1979.
139. SCOTT, JAMES H., JR. AND ROTHMAN, MITCHELL P. "The Effect of an Introductory Economics Course on Student Political Attitudes," *J. Econ. Ed.*, Spring 1975, 6(2), pp. 107-12.
140. SIEGFRIED, JOHN J. "A Reply to the Comment of Professors Battalio, Hulett, and Kagel on 'The Publishing of Economic Papers and Its Impact on Graduate Faculty Ratings, 1960-1969,'" *J. Econ. Lit.*, March 1973, 11(1), pp. 71-73.
141. ———. "Is Teaching the Best Way to Learn? An Evaluation of the Benefits and Costs to Undergraduate Student Proctors in Elementary Economics," *Southern Econ. J.*, Jan. 1977, 43(3), pp. 1394-1400.
142. ———. "Male-Female Differences in Economics Education: A Survey," *J. Econ. Ed.*, Spring 1979, 10(2).
143. ——— AND STRAND, STEPHEN H. "An Evaluation of the Vanderbilt-JCEE Experimental PSI Course in Elementary Economics," *J. Econ. Ed.*, Fall 1976, 8(1), pp. 9-26.
144. ——— AND STRAND, STEPHEN H. "Sex and the Economics Student," *Rev. Econ. Statist.*, May 1977, 59(2), pp. 247-49.
145. ——— AND SWEENEY, GEORGE H. "Bias in Economics Education Research from Random and Voluntary Selection into Experimental and Control Groups." Unpublished manuscript, Vanderbilt University, 1979.
146. ——— AND WHITE, KENNETH J. "Financial Rewards to Research and Teaching: A Case Study of Academic Economists," *Amer. Econ. Rev.*, May 1973, 63(2), pp. 309-15.
147. ——— AND WHITE, KENNETH J. "Teaching Ability as a Determinant of Faculty Salaries," *J. Econ. Ed.*, Spring 1978, 9(2), pp. 130-32.
148. SKINNER, B. F. AND HOLLAND, JAMES G. *The analysis of behavior: A program for self-instruction*. New York: McGraw-Hill, 1961.
149. SLOANE, PETER E. "The Relationship of Performance to Instruction and Student Attitudes," in WELSH, ARTHUR L., ed. [173, 1972], pp. 70-83.
150. SOPER, JOHN C. "Programmed Instruction in Large-Lecture Courses," *J. Econ. Ed.*, Spring 1973, 4(2), pp. 125-29.
151. ———. "Soft Research on a Hard Subject: Student Evaluations Reconsidered," *J. Econ. Ed.*, Fall 1973, 5(1), pp. 22-26.
152. ———. "Computer-Assisted Instruction in Economics: A Survey," *J. Econ. Ed.*, Fall 1974, 6(1), pp. 5-28.
153. ———. "Second Generation Research in Economic Education: Problems of Specification and Interdependence," *J. Econ. Ed.*, Fall 1976, 8(1), pp. 40-48.
154. ———. *Test of Economic Literacy: Discussion guide and rationale*. New York: Joint Council on Economic Education, 1979.
155. ——— AND THORNTON, RICHARD M. "Self-Paced Economics Instruction: A Large-Scale Disaggregated Evaluation," *J. Econ. Ed.*, Spring 1976, 7(2), pp. 81-91.

156. SOSIN, KIM AND MCCONNELL, CAMPBELL R. "Student Perceptions of Income Distribution." Unpublished manuscript, University of Nebraska, circa 1978.
157. SPECTOR, LEE AND MAZZEO, MICHAEL. "Probit Analysis and Economic Education." Unpublished manuscript, State University College at Buffalo, 1978.
158. STIGLER, GEORGE J. "The Politics of Political Economists," *Quart. J. Econ.*, Nov. 1959, 73(4), pp. 522-32.
159. ———. "Elementary Economic Education," *Amer. Econ. Rev.*, May 1963, 53(2), pp. 653-59.
160. SWAN, CRAIG. "Comments on the Problems of Specification and Interdependence in a Set of Learning Equations," *J. Econ. Ed.*, Spring 1978, 9(2), pp. 81-86.
161. SWARTZ, T. R.; BONELLO, F. J. AND DAVISSON, W. I. "The Use of TUCE May Lead to Misunderstanding Cognitive Achievement." Unpublished paper presented at the June 1977 meetings of the Western Economic Association and the November 1977 meetings of the National Association of Affiliated Economic Education Directors.
162. TAYLOR, HORACE, ed. "The Teaching of Undergraduate Economics: Report of the Committee on the Undergraduate Teaching of Economics and Training of Economists," *Amer. Econ. Rev.*, Dec. 1950, Supplement, Part 2, 40(5).
163. THOMPSON, FRED A. "The Interaction of Cognition and Affect: The Issue of Free Trade," *J. Econ. Ed.*, Spring 1973, 4(2), pp. 111-15.
164. TIEMANN, PHILIP W.; PADEN, DONALD W. AND MCINTYRE, CHARLES J. *An application of the principles of programmed instruction to a televised course in college economics*. Washington, D.C.: U.S. Department of Health, Education and Welfare, 1966.
165. TUCKMAN, BARBARA AND TUCKMAN, HOWARD. *Toward a more effective economic principles class: The Florida State University experience*. *J. Econ. Ed.*, Spring 1975, Special Issue Number 3. New York: Joint Council on Economic Education, 1975.
166. TUCKMAN, HOWARD P. "Teacher Effectiveness and Student Performance," *J. Econ. Ed.*, Fall 1975, 7(1), pp. 34-39.
167. ———. *Publication, teaching, and the academic reward structure*. Lexington, Mass.: Heath, Lexington Books, 1976.
168. WEHRS, WILLIAM. "Incentive Structure and the TUCE," *J. Econ. Ed.*, Spring 1978, 9(2), pp. 107-10.
169. WEIDENAAR, DENNIS J. "The Effect of Programmed Instructional Materials and Other Factors on Performance in the Introductory Economics Course," in WELSH, ARTHUR L., ed. [173, 1972], pp. 118-32.
170. ——— AND DODSON, JOE A., JR. "The Effectiveness of Economics Instruction in Two-Year Colleges," *J. Econ. Ed.*, Fall 1972, 4(1), pp. 5-12.
171. WEISBROD, BURTON A. "Research on Economic Education: Is It Asking the Right Questions?" *Amer. Econ. Rev.*, May 1979, 69(2), pp. 14-21.
172. WELLS, STUART. "The Effectiveness of Economics Education: A Survey of Research Findings," Faculty of Commerce, University of British Columbia, Working Paper Number 331, November 1974.
173. WELSH, ARTHUR L., ed. *Research papers in economic education*. New York: Joint Council on Economic Education, 1972.
174. ——— AND FELS, RENDIGS. "Performance on the New Test of Understanding in College Economics,"

Siegfried and Fels: Teaching College Economics

- Amer. Econ. Rev.*, May 1969, 59(2), pp. 224-29.
175. WENTWORTH, DONALD R.; HANSEN, W. LEE AND HAWKE, SHARRYL H. *Perspectives on economic education*. New York: Joint Council on Economic Education, 1977.
176. ——— AND LEWIS, DARRELL R. "An Evaluation of the Use of the Marketplace Game in Junior College Economics," *J. Econ. Ed.*, Spring 1975, 6(2), pp. 113-19.
177. WHITNEY, SIMON N. "Measuring the Success of the Elementary Course," *Amer. Econ. Rev.*, March 1960, 50(1), pp. 159-69.
178. YATES, JUDITH. "Research in Economics Education: Are Our Horizons Too Narrow?" *J. Econ. Ed.*, Fall 1978, 10(1), pp. 12-17.
179. ZWEIG, MICHAEL. "Teaching Radical Political Economics in the Introductory Course," *Amer. Econ. Rev.*, May 1972, 62(2), pp. 434-38.

Male-Female Differences in Economic Education: a Survey

John J. Siegfried

There have been numerous efforts to examine the relationship between gender and student performance in the college-level introductory economics course. Most such studies have examined the association of gender with student learning as a by-product of their primary objective.¹ Student gender is usually added as a control variable in statistical studies associating some experimental teaching technique with student learning. For this reason, many of the studies which evaluate the association of sex with student learning suffer from methodological deficiencies.

There also have been some studies of the association of gender with student performance in economics in primary grades and in high school. Finally, there is a little evidence concerning gender differences in understanding economics among college graduates.

This survey first examines the hypotheses that are commonly used to justify the inclusion of a binary variable for gender in models of the "production function" for learning. Second, the distinction between *understanding* and *learning* economics is made, and its implications for the empirical studies are described. Third, the many empirical studies examining the association of gender with economic education are summarized and evaluated. The effect of gender on learning and understanding is distinguished and the question of when gender differentials appear is addressed. Fourth, gender differences in the effectiveness of alternative teaching techniques is discussed. The final section reports differences between men and women students in their enjoyment of and interest in economics.

Hypotheses

The standard hypothesis underlying most empirical tests of the effect of gender on economic learning or understanding is that female students have grown up in a cultural environment in which girls are not supposed to like business and thus have a disadvantage in business or economics courses. More recently, Garron has argued that the difference between male and female performance in learning spatial and numerical skills, which are related to understanding economics, is chromosome-linked.

MacDowell, Senn, and Soper group the sociocultural explanations into four different, though related, categories. First, they report that the psychological literature consistently finds that identity issues in adolescent women tend to focus on

John J. Siegfried is associate professor of economics at Vanderbilt University. Phillip Saunders and Thomas Zak provided helpful comments on an earlier version. Jack Fiedler and James M. Lewek assisted in the research.

their search for a husband, and many think and act as though they will be rejected by boys if they depart far from adolescent stereotypes, one of which is that business is a man's world. Second, there apparently is evidence that sex stereotypes are stronger in higher-income families. Since lower-income families devote a greater share of their available income to sending male offspring to college (because college support is rationed to sons before daughters), there will be a greater than average proportion of women in college who subscribe to sex stereotypes. Third, young females are significantly more dependent than males. It is possible that economics courses are taught in ways that penalize dependency, for example, in large lecture formats. Thus teaching techniques may be better suited to males than females. Finally, differences in maturation rates may explain differences in economics skills. The psychological literature finds that people who mature earlier have higher verbal learning rates while spatial skills are less developed for any given age. Since women generally mature earlier than men, we would expect women to have a comparative advantage in verbal skills while men obtained a comparative advantage in spatial (economics?) skills.² Ladd adds a fifth possibility: that teacher attitudes and learning material, especially in precollege education, may be at fault.

There is, of course, no reason to single out any one cause of the differences in economics mastery between men and women. Several or all of the hypotheses may be working simultaneously.

Understanding Versus Learning Economics

The level of understanding is the stock of knowledge about economics and business at a point in time. Learning is the flow of new knowledge that occurs over a period of time.

If there is a concern about the differences in economic understanding between adult males and females, and one desires to find out what actions might be taken to reduce these differences (presumably by improving the cognitive levels of women rather than reducing those of men), then determining the age at which the differentials appear is important. For example, do the differences in understanding of business and economic concepts grow slowly over the years, or is there a pattern of no differences until a certain age after which differences appear and then remain stable thereafter. If the latter pattern characterizes the development of differences in understanding, then we could focus corrective efforts on people during a certain stage of development. On the other hand, if the differentials grow slowly, the cost of reducing the differentials is likely to be greater since action would have to be taken over a broader group of women, of various ages.

Most of the studies of gender and student performance in economics fail to consider explicitly whether it is the stock or flow of knowledge that they expect gender to influence. The design of most experiments that include a binary variable for gender is aimed primarily at evaluating some alternative teaching technique; so, little attention is devoted to whether the impact of gender on performance is specified sensibly. Consequently some studies measure student understanding at a point in time and other studies measure learning of economics over a period of time (usually during a course). This variety provides an opportunity to distinguish the effect of gender on the stock of economic understanding from the effect on flow.

Studies that examine the stock of knowledge at a point in time generally report correlations between student performance on a final examination and student gender. Studies that assess the relationship between gender and the flow of knowledge either

use a "value-added" measure of performance, or include a pretest control variable in their model before correlating a post-test performance measure with gender.

The Empirical Evidence

It is difficult to summarize and compare the various studies of gender and economic education because they vary considerably in research methodology, sample size, and unit of observation, and frequently they confuse the hypotheses. In addition, different levels of statistical significance are used, with the result that a finding which would be statistically significant at the .05 level may be declared insignificant if the significance criterion is .01. Rather than try to unravel these differences in statistical methodology, we have simply based significance conclusions on the criteria employed in the particular studies. Therefore, the summary provides only a rough indication of the findings. Finally, although a count of the studies which find positive or no effects of gender on economic education may be illuminating, one must recognize that the quality of the research varies, and there is no compelling reason to assign equal weights to each of the studies.

Two-thirds of the studies that related the *level of understanding* with gender found that men performed statistically significantly better than women. Only one-third of the studies that examined the *flow of students' learning* during (mostly college) courses found that men did statistically significantly better. Although there is a substantial division of opinion, in general the empirical research seems to suggest that by the time people reach college age, men are significantly ahead of women in understanding economics, but both sexes are progressing at equivalent rates; thus the gap would appear to be stable by that time. These results are consistent with Buckles and Welsh's findings that there is very little change in the rank of students in the principles course from the beginning to the end.

Data on differences in understanding economics between the sexes by level of schooling can reveal the time when such differences develop. In a study of elementary-level students, Davison and Kilgore found no sex-related differences in the stock of economics understanding. Ladd surveys evidence that concurs with the Davison-Kilgore conclusion. By the time students graduate from high school, differences begin to appear. Highsmith in a study of grades 7-12, Moyer and Paden in a study of high school students, and Thornton and Vredeveld, also examining high school students, all found a statistically significant advantage for males in the level of understanding of economics. Becker, Helmberger, and Thompson report contradictory evidence while evaluating Project DEEP's effect on high school students' understanding of economics. MacDowell, Senn, and Soper have done the most extensive study of the effect of gender on high school students' performance in economics. In an evaluation of the World of Work Economic Education (WOWEE) Project, they gathered test scores for about two thousand Illinois junior and senior high school students on the nationally normed *Junior High School Test of Economics*. They regressed post-WOWEE test scores on pretest scores, a binary variable for students' gender, a binary variable for teacher's gender, and a binary variable which was 1 if the student's and teacher's gender were the same, and zero otherwise. Their results indicate no difference in post-test scores on the basis of students' gender. The students of male teachers did significantly better. The student-teacher interaction variable was nonsignificant. MacDowell-Senn-Soper conclude that gender differences have not yet surfaced by about age 15, the typical age of students in their sample. However, this conclusion seems unwarranted because they control for pretest scores in their analysis. The measured effect of gender in their study is on the flow of

knowledge during the World of Work Project. If males were already ahead of females by the time the project commenced, the empirical results would not reveal the difference. Thus, their study appears to reveal no gender-related differences in student learning of economics in secondary schools.

The studies all show that the difference in economics understanding between males and females has already developed by high school. However, it is impossible to quantify the magnitude of the effect and compare it to differences at the college level (in order to determine if further widening of the gap occurs at the transition from high school to college) because of the absence of a standardized measurement instrument that would be appropriate for both high school and college students.

There are many studies on college students, perhaps because of the relatively low cost of acquiring data. For junior college, Thompson finds no difference in *levels of understanding* of economics. Wentworth and Lewis, Weidenaar and Dodson, and Lewis, Wentworth, and Orvis find no relationship between *learning* introductory economics and students' gender.

For students in the principles-of-economics course in four-year colleges in the United States, many studies reveal *no difference in learning* of economics during the course: Buckles and McMahon; Kelley (1975); Paden and Moyer; Lewis and Dahl (1972a); Weidenaar; Gery; Ramsett, Johnson, and Adams; Siegfried and Strand; and Elliott, Ireland, and Cannon. In contrast, there are relatively few studies which report that there is a *statistically significant difference in learning* of economics during the principles course on the basis of gender: Crowley and Wilton; Sloane; Tuckman; Soper and Thornton; and Soper (1973, 1976). There exist carefully done studies with relatively large samples on both sides of this issue, but the preponderance of the evidence on student *learning* and gender in introductory economics suggests *no difference*.

The findings regarding gender-related differences in *level of understanding* in the college principles-of-economics course in the United States contrast with the absence of a gender difference in learning. Those studies finding that men do significantly better include Bach and Saunders; Bolch and Fels; Allison (1976a, 1976b, 1977); Marston and Lyon; Soper and Thornton; Lewis and Orvis; Soper (1973); Chizmar, Hiebert, and McCarney; Clauretje and Johnson; Siegfried and Strand; and Saunders (1975). In contrast, no gender difference even on *absolute levels of performance* at the principles level was discovered by Paden, Dalgaard, and Barr; Emery and Enger; Lewis and Dahl (1972b); Danielsen and Stauffer; Marston, Lyon, and Knight; and Morgan and Vasche. Here the weight of methodological superiority and sample size belongs to those studies which have found a statistically significant advantage for men in absolute performance.³

Although Harbury and Szreter report no gender differences in the level of understanding of economics among U.K. college students, in a study of 4,700 U.K. college students at thirty-seven universities, Attiyeh and Lumsden (1971, 1972) report that men do better on measures of both understanding and learning. Palmer, Carliner, and Romer found no learning difference between men and women in a Canadian introductory course.

The tests at the principles-of-economics level were based on samples for all types of colleges and universities (large and small, private and public, northern and southern), using various evaluation instruments (TUCE, TEU, TEC, and CLEP), and have been parts of evaluations of all types of teaching techniques (programmed learning, self-paced instruction, televised lectures, computer-assisted instruction).

There is evidence that male-female differences in understanding persist beyond

the introductory course, and may even grow. Saunders and Bach and Bach and Saunders (1965, 1966) assessed the level of understanding of seniors in college, high school teachers who watched the television program, "The American Economy," and a large sample of graduates of U.S. colleges. They found that men did better on almost all of their tests of understanding of economics. The most comprehensive evidence available on the persistent differences in understanding between female and male adults is from the Bach and Saunders (1966) and the Saunders (1973) studies. Bach and Saunders found that men who graduated from prestige schools, large universities, teachers colleges, and "other schools" did significantly better on the TEU than women with similar educational backgrounds. The sample sizes were 103, 503, 1,251, and 1,945, respectively. The diversity of schools represented in the sample lends considerable confidence to their findings. Likewise, Saunders' regressions with 1,220 sophomore, 955 senior, and 1,257 alumni respondents from twenty-five carefully selected schools found a significantly better performance of males than females on a hybrid TUCE for each group. Since there was no pretest control or gap-closing measure of achievement, Saunders' study confirms the superior level of understanding of economics by males, but not their level of learning. The difference showed up at the sophomore level and increased slightly through the progression to senior and alumni status.

More recently, Kohen and Kipps, in a study of 59 students at James Madison University, found no difference in the TUCE scores between men and women who were beginning an intermediate microeconomics course, after controlling for achievement levels at the end of introductory economics, time elapsed since taking introductory economics, overall grade point average, and number of credit-hours in business-economics-mathematics courses. A test for interaction between gender and elapsed time would have revealed any differences in the rate of depreciation of economic knowledge, but was not done.

Bach and Saunders (1966) is the only study surveyed which found a group with a statistically significant *advantage* for women; it was for people who had graduated from fifty liberal arts colleges. The sample size was 138. However, Rothman and Scott find no difference between men and women on a post-TUCE examination when women were at a significant disadvantage at the beginning of the course (measured by pre-TUCE), implying greater learning by women during the course. In a study of what students learn from the economics major at Harvard after their introductory economics course, Hartman found no differences between men and women.

Probably the greatest insight into the reasons for differences in performance of men and women on economics achievement tests is found in Allison's (1977) investigation of the educational production process in introductory economics at Harvard. Using a three-equation, simultaneous nonlinear equations model of the learning process, she found that there was no gap between men and women in understanding economics by November of the fall semester, but by January men performed slightly better, and by June (of the second semester) men performed significantly better. A separate equation in the model confirmed that this difference was not due to a greater effort being exerted by men.

The explanation, according to Allison, was uncovered in separate estimates of the model for men and women, which revealed that the marginal product of hours spent studying economics was greater for men than for women. When the year-long course began men had about a two to one advantage in marginal product of effort, and by the end of the course this gap had widened. Consequently, Allison developed the following account of the source of differential performance: "Women enter the

course with less "skill" in learning analytic material—less practice in that peculiar intellectual exercise of model building. Thus, learning analytic material comes slowly; much of each hour spent studying is misdirected. Consequently, at the end of the semester they have learned relatively little economics per unit of time, and even less economic theory. Nor have they learned on the average analytical material as well as their male counterparts. Thus, in the second semester women are, given the cumulative nature of economics, doubly disadvantaged. Since (as revealed by other evidence in the study) they are more sensitive to grades than their male counterparts, they do not reduce their effort during the second semester. But as indicated in the second semester enjoyment equation, they ultimately find the experience relatively unsatisfying" (1977, pp. 40-41).

Allison's account is consistent with evidence on the difference in performance between men and women on particular types of achievement examination questions. She found a much greater disadvantage for women on theory questions than on other types of questions. Saunders (1975) found that the positive association of male sex with test performance (level of understanding) on the hybrid TUCE tended to be stronger on simple application and complex application questions than on recognition and understanding questions, although the differences were not significant (and there is one exception). Elliott, Ireland, and Cannon also found the disadvantage of females to be greater on application than recognition and understanding questions in a study that measured learning during the introductory course. Finally, Allison's account of the time pattern of the difference in test performance is in accord with her explanation that its source is based in analytical skills differences,⁴ since the insignificant gender difference in performance by November of the first semester is more likely to reflect achievement on recognition and understanding skills vis-à-vis the commonly greater analytical content included in examinations as the course progresses.⁵

Interactions Between Teaching Techniques, Ability, and Sex Differences in Learning Economics

Chizmar, Hiebert, and McCarney ran separate regressions to explain the level of economic understanding of students at Illinois State University in a computer-assisted instruction class and a class without CAI. They found no gender-related differences in the level of economic understanding on the basis of the method of teaching. In contrast, Allison (1977) and Siegfried and Strand found that women taking the principles-of-economics course at Harvard and Vanderbilt did relatively better in a self-paced-instruction (SPI) class than in a mass lecture. This finding is consistent with the theory that women respond better to environments that accommodate dependence: in an SPI course, tutors are assigned to each student and work closely with the individual. However, the evidence is also consistent with Allison's account, namely, that the difference in analytical skills is the fundamental source of the gap. Drill in analytical materials is generally more intense in SPI courses, thereby focusing instruction on the weak area of preparation for most women.

There are several interesting findings that go beyond the simple average male-female difference in performance on economics tests. For example, Sloane found that men did better overall using a learning criterion in his principles-of-economics course, but the gender difference disappeared for the top and bottom quartiles of the class. Apparently it is manifested almost entirely in the middle range of the class (ranges determined by pre-TUCE). Sloane also found that women did relatively better with the more inexperienced instructor than did men. Allison (1977) also analyzed the effect of gender on achievement separately for high- and

low-achieving students. She found that the female disadvantage was more pronounced for the bottom quarter of achieving students. Allison does not report similar subregressions for the middle two quarters. Consequently, although her findings do not appear to be similar to Sloane's finding regarding a nonlinear effect of achievement levels on sex differences in understanding economics, a conclusive determination cannot be made.

Sex Differences in Attitudes toward Economics

Allison (1976a, 1977) reports that women showed a statistically significantly greater level of enjoyment of the principles-of-economics course at Harvard even though they understood less economics. In spite of their greater enjoyment and lower achievement, there were no apparent differences in work effort related to student gender. Kelley (1972) also found that women tend to like the principles-of-economics course better, but he found no difference in ratings of the professor on the basis of student gender. Wentworth and Lewis and Paden and Moyer, however, report no significant male-female differences in attitudes toward economic instruction.

In his study of the lasting effects of economic instruction, Saunders (1973) found greater interest in economics (as self-reported on a questionnaire) among men than women for sophomores, seniors, and alumni who had not had any college economics. However, among students who had had an introductory course, there was no significant difference in levels of interest between men and women. This suggests that interest differences between the sexes are manifested in course selection rather than being an important determinant of variations in performance during the course.

Tuckman found that males are more likely to continue their economic training by taking further courses. Allison (1976a), however, found that gender was useless as a predictor of re-enrollment in economics courses.

Mann and Fushfeld found that men had both a greater level of sophistication of attitudes toward economics and gained a little more attitude sophistication during the course than did women. However, their measure of attitude sophistication has been criticized by Rothman and Scott, for, among other reasons, being a measure of political liberal values. More recently, Sosin and McConnell examined the effect of an introductory macroeconomics course on student attitudes toward the distribution of income. They found that, in general, the course led to a statistically significant move in attitudes away from conservatism and toward an egalitarian attitude about income distribution. They hypothesized that women would experience a greater shift in their attitudes toward an egalitarian distribution of income because they might anticipate a work-life cycle involving periods of dependence upon intrahousehold transfers of income. The empirical results supported their hypothesis.

Davisson and Bonello, in a large study of computer-assisted instruction at Notre Dame, found comparatively little difference between men and women in attitudes toward economic institutions, problems, and policies. They asked students to classify, for example, government spending deficits, government controls of wages and prices, poverty, inflation, increasing the money supply, labor unions, big business, market mechanisms, etc., as good or bad, inevitable or controllable, effective or ineffective, etc.

Conclusion

The scant evidence on learning and understanding economics at the elementary school level indicates few differences between the sexes. However, by the high school years gaps appear to develop. Differences in understanding seem to persist

through the college years, but there does not appear to be any widening of the gap. Most of the research from which these conclusions have been drawn was designed for other purposes, and consequently is not satisfactory enough to resolve the issue definitively.

FOOTNOTES

1. The primary objective is usually to assess a particular teaching method. See Siegfried and Fels for an extensive catalog and critique of these studies.
2. The puzzle surrounding this argument, however, is that most studies of achievement in introductory economics find that verbal SAT rather than quantitative SAT is more important in explaining success (see Fels and Siegfried).
3. In a study of the economic statistics course, Cohn found no significant difference between the performance of men and women in his course, although the sample of women was relatively small.
4. In a recent survey of male-female differences in precollege economics education, Ladd cites evidence that there are no differences in analytic ability or concept mastery between men and women. Apparently recent work in psychology argues that the commonly perceived difference is confined more narrowly to differences in visual spatial skills than general analytic ability. Ladd points to differences in quantitative and verbal abilities as a factor that explains male-female performance differences. "If economics at the precollege level requires more quantitative ability than verbal ability (a debatable issue), however, some differential learning ability for economics may exist" (p. 147). However, this is contrary to the relatively greater impact of SAT verbal scores than SAT quantitative scores on cognitive achievement in introductory economics courses (see Siegfried and Fels).
5. Another possible explanation for the observed difference between men and women in economics achievement is sample self-selection bias. For example, if general tastes for learning economics vary systematically by sex, and the appropriate skill distribution for men and women is similar, the difference in observed performance could be due to comparing the tail (upper, presumably) of one distribution, likely female, with a broader cross section of the other distribution, male. Such a possibility would normally lead one to expect a bias toward showing that women do better than men. That the studies almost all show either no effect or worse performance by women is not necessarily inconsistent with this idea. Indeed, this bias could well mean that the true disadvantage of women in learning and understanding economics is understated by the reported empirical tests because the tests are biased in favor of showing an advantage for women, due to the selectivity of women who elect economics. One way of learning more about this process would be to explore the effect of interaction variables (between the usual binary variable for gender and some taste variable, say major) in the regression analyses. To my knowledge this has never been done.

REFERENCES

- Elisabeth K. Allison, "Three Years of Self-Paced Teaching in Introductory Economics at Harvard," *American Economic Review*, 66 (May 1976a), pp. 222-228.
- , "The Use of Video in Economic Education," *The Journal of Economic Education*, 8 (Fall 1976b), pp. 27-36.
- , "Educational Production Function for an Introductory Economics Course," *Harvard University Discussion Paper*, 345 (April 1977).
- Richard Attiyeh and Keith G. Lumsden, "University Students' Initial Understanding of Economics: The Contribution of the A Level Economics Course and of Other Factors," *Economica*, 38 (February 1971), pp. 81-97.
- , "Some Modern Myths in Teaching Economics: The U.K. Experience," *American Economic Review*, 62 (May 1972), pp. 429-433.
- G. L. Bach and Phillip Saunders, "Economic Education: Aspirations and Achievements,"

- American Economic Review*, 55 (June 1965), pp. 329-356.
- , "Lasting Effects of Economics Courses at Different Types of Institutions," *American Economic Review*, 56 (June 1966), pp. 505-511.
- William E. Becker, Jr., John D. Helmberger, and Jerry L. Thompson, "An Evaluation Project Given Limited Data," *JEE*, 6 (Spring 1975), pp. 120-125.
- Ben W. Bolch and Rendigs Fels, "A Note on Sex and Economic Education," *JEE*, 6 (Fall 1974), pp. 64-67.
- Stephen G. Buckles and Marshall E. McMahon, "Further Evidence on the Value of Lectures in Elementary Economics," *JEE*, (Spring 1971), pp. 138-141.
- Stephen G. Buckles and A. L. Welsh, "The Use of Validated Tests in Teaching and Research." In A. L. Welsh, ed., *Research Papers in Economic Education*. New York: Joint Council on Economic Education, 1972.
- John F. Chizmar, L. Dean Hiebert, and Bernard J. McCarney, "Assessing the Impact of an Instructional Innovation on Achievement Differentials: The Case of Computer-Assisted Instruction," *JEE*, 9 (Fall 1977), pp. 42-46.
- Terrence M. Clauretie and William E. Johnson, "Factors Affecting Student Performance in Principles of Economics," *JEE*, 6 (Spring 1975), pp. 132-134.
- Elchanan Cohn, "Students' Characteristics and Performance in Economic Statistics," *JEE*, 3 (Spring 1972), pp. 106-111.
- Ronald W. Crowley and David A. Wilton, "A Preliminary Report on the Efficiency of Introductory Economics Courses," *JEE*, 5 (Spring 1974), pp. 103-108.
- Albert L. Danielsen and A. J. Stauffer, "A Television Experiment in College Economics," *JEE*, 3 (Spring 1972) pp. 101-105.
- Donald G. Davison and John H. Kilgore, "A Model for Evaluating the Effectiveness of Economic Education in Primary Grades," *JEE*, 3 (Fall 1971), pp. 17-25.
- William I. Davisson and Frank J. Bonello, *Computer-Assisted Instruction in Economic Education*. Notre Dame, Ind.: University of Notre Dame Press, 1976.
- Ralph D. Elliott, M. Edwin Ireland, and Teresa S. Cannon, "An Analysis of the Marginal Products of the One- and Two-Semester Introductory Economics Courses," *JEE*, 10 (Fall 1978), pp. 18-25.
- E. David Emery and Thomas P. Enger, "Computer Gaming and Learning in an Introductory Economics Course," *JEE*, 3 (Spring 1972), pp. 77-85.
- C. Garron, "Sex-Linked, Recessive Inheritance of Spatial and Numerical Abilities, and Turner's Syndrome," *Psychological Review*, 77 (May 1970).
- Frank W. Gery, "Is There a Ceiling Effect to the Test of Understanding in College Economics?" In A. L. Welsh, ed., *Research Papers in Economic Education*. New York: Joint Council on Economic Education, 1972.
- C. D. Harbury and R. Szepter, "The Value of Prior Experience of Economics for University Students," *JEE*, 2 (Fall 1970), pp. 56-62.
- David G. Hartman, "What Do Economics Majors Learn?" *American Economic Review*, 68 (May 1978), pp. 17-22.
- Robert Highsmith, "A Study to Measure the Impact of In-Service Institutes on the Students of Teachers Who Have Participated," *JEE*, 5 (Spring 1974), pp. 77-81.
- Allen C. Kelley, "Uses and Abuses of Course Evaluations as Measures of Educational Output," *JEE*, 4 (Fall 1972), pp. 13-18.
- , "The Student as a Utility Maximizer," *JEE*, 6 (Spring 1975), pp. 82-92.
- Andrew I. Kohen and Paul H. Kipps, "Factors Determining Student Retention of Economic Knowledge After the Principles-of-Microeconomics Course," *JEE*, 10 (Spring 1979), pp. 38-48.
- Helen F. Ladd, "Male-Female Differences in Precollege Economic Education." In Donald R.

- Wentworth, W. Lee Hansen, and Sharryl H. Hawke, eds., *Perspectives on Economic Education*. New York: Joint Council on Economic Education, 1977.
- Darrell R. Lewis and Tor Dahl, "Critical Thinking Skills in the Principles Course: An Experiment." In A. L. Welsh, ed., *Research Papers in Economic Education*. New York: Joint Council on Economic Education, 1972a.
- , "Factors Influencing Performance in the Principles Course Revisited." In A. L. Welsh, ed., *Research Papers in Economic Education*. New York: Joint Council on Economic Education, 1972b.
- Darrell R. Lewis and Charles C. Orvis, "A Training System for Graduate Student Instructors of Introductory Economics at the University of Minnesota," *JEE*, 5 (Fall 1973), pp. 38-46.
- Darrell R. Lewis, Donald R. Wentworth, and Charles C. Orvis, "Economics in the Junior Colleges: Terminal or Transfer?" *JEE*, 4 (Spring 1973), pp. 100-110.
- Michael A. MacDowell, Peter R. Senn, and John C. Soper, "Does Sex Really Matter?" *JEE*, 9 (Fall 1977), pp. 28-33.
- William R. Mann and Daniel R. Fوسفeld, "Attitude Sophistication and Effective Teaching in Economics," *JEE*, 1 (Spring 1970), pp. 111-129.
- Glenn F. Marston and Kenneth S. Lyon, "Learning and Attitude Change of Students Subjected to a National Income Simulation Game: Some Further Evidence," *JEE*, 7 (Fall 1975), pp. 20-27.
- Glenn F. Marston, Kenneth S. Lyon, and Richard Knight, "Learning and Attitude Change of Students Subjected to a National Income Simulation Game." In A. L. Welsh, ed., *Research Papers in Economic Education*. New York: Joint Council on Economic Education, 1972.
- W. Douglas Morgan and Jon David Vasche, "An Education Production Function Approach to Teaching Effectiveness and Evaluation," *JEE*, 8 (Spring 1978), pp. 123-126.
- M. E. Moyer and D. W. Paden, "On The Efficiency of the High School Economics Course," *American Economic Review*, 58 (September 1968), pp. 870-877.
- Donald W. Paden, Bruce R. Dalggaard, and Michael D. Barr, "A Decade of Computer-Assisted Instruction," *JEE*, 9 (Fall 1977), pp. 14-20.
- Donald W. Paden and M. Eugene Moyer, "The Relative Effectiveness of Three Methods of Teaching Principles of Economics," *JEE*, 1 (Fall 1969), pp. 33-45.
- John Palmer, Geoffrey Carliner, and Thomas Romer, "Does High School Economics Help?" *JEE*, 10 (Spring 1979).
- David E. Ramsett, Jerry D. Johnson, and Curtis Adams, "Some Evidence on the Value of Instructors in Teaching Economic Principles," *JEE*, 5 (Fall 1973), pp. 57-62.
- Mitchell P. Rothman and James H. Scott, "Political Opinions and the TUCE," *JEE*, 4 (Spring 1973), pp. 116-124.
- Phillip Saunders, "The Lasting Effects of Elementary Economics Courses," *Final Report*, NSF Grant GY-7208 (1973).
- , "Experimental Course Development in Introductory Economics at Indiana University," *JEE*, 7 (Fall 1975), Special Issue.
- Phillip Saunders and G. L. Bach, "The Lasting Effects of an Introductory Economics Course: An Exploratory Study," *JEE*, 1 (Spring 1970), pp. 143-149.
- John J. Siegfried and Rendigs Fels, "Research on Teaching College Economics: A Survey," *Journal of Economic Literature*, 16 (September 1979).
- John J. Siegfried and Stephen Strand, "Sex and the Economics Student," *Review of Economics and Statistics*, 59 (May 1977), pp. 247-249.
- Peter E. Sloane, "The Relationship of Performance to Instruction and Student Attitudes." In

- A. L. Welsh, ed., *Research Papers in Economic Education*. New York: Joint Council on Economic Education, 1972.
- John C. Soper, "Second Generation Research in Economic Education: Problems of Specification and Interdependence," *JEE*, 8 (Fall 1976), pp. 40-48.
- , "Programmed Instruction in Large-Lecture Courses," *JEE*, 4 (Spring 1973), pp. 125-129.
- John C. Soper and Richard M. Thornton, "Self-Paced Economic Instruction: A Large-Scale Disaggregated Evaluation," *JEE*, 7 (Spring 1976), pp. 81-91.
- Kim Sosin and Campbell R. McConnell, "Student Perceptions on Income Distribution," University of Nebraska-Lincoln, no date. MS. only.
- Fred A. Thompson, "Problems and Prospects of Economics Education in Community Junior Colleges," *JEE*, 2 (Fall 1970), pp. 31-38.
- Daniel L. Thornton and George M. Vredevelt, "In-Service Education and Its Effect on Secondary Students: A New Approach," *JEE*, 8 (Spring 1977), pp. 93-99.
- Howard P. Tuckman, "Teacher Effectiveness and Student Performance," *JEE*, 7 (Fall 1975), pp. 34-39.
- Dennis J. Weidenaar, "The Effect of Programmed Instructional Materials and Other Factors on Performance in the Introductory Economics Course." In A. L. Welsh, ed., *Research Papers in Economic Education*. New York: Joint Council on Economic Education, 1972.
- Dennis J. Weidenaar and Joe A. Dodson, Jr., "The Effectiveness of Economics Instruction in Two-Year Colleges," *JEE*, 4 (Fall 1972), pp. 5-12.
- Donald R. Wentworth and Darrell R. Lewis, "An Evaluation of the Use of the Marketplace Game in Junior College Economics," *JEE*, 6 (Spring 1975), pp. 113-119.

II

Critiques of Research Methodology

ECONOMIC EDUCATION RESEARCH: ISSUES AND ANSWERS

Research on Economic Education:
Is It Asking the Right Questions?

By BURTON A. WEISBROD*

The division of responsibility between the two papers at this session is a fascinating one. One author has agreed to examine the questions that are being asked in the economics education literature, and the other to examine the answers!

As is so often the case, however, the underlying assumption of separability does not hold. A research question is not a "good" or "bad" question independent of the quality of the answers it is likely to generate. An "exciting" question that is unlikely to yield an answer of substantial value is *not* a good question. Research is a production process in which something called "useful knowledge" is the output. The inputs to this process include both the specification of questions that are important—in the sense that the answers would have great expected value—and the marshaling of resources (i.e., the incurring of costs) to answer the questions.

If the costs of answering all research questions were equal, or were random with respect to the significance of the question, then the separability of the decisions on question *specification* and on question *answering* would be justified. What we probably confront, however, is a less fortuitous set of conditions in which the questions that are most valuable to answer are also the most costly (i.e., difficult). The issue I have been asked to deal with—whether research on economic educa-

tion is asking the right questions—thus involves implicitly the performance of a benefit-cost analysis on project selection in the area of economic education research. An evaluation is needed of (a) the expected benefits (more precisely, the probability distribution of benefits conditional on answers of various quality), and (b) the expected costs of obtaining answers of each quality.

It is possible, of course, that a particular research question may be a good one in the efficiency sense that the expected costs of researching it are less than what the expected benefits would be *if* the resources devoted to the research were used as productively as possible; yet if the resources were not used so productively, it might fail the allocative efficiency test. Thus, a question could be *potentially* efficient to research but *actually* inefficient. In any event, the "best" questions to research are those for which the excess of the value of the expected answers (benefits) over the expected costs of the research are maximized. Deciding which are the "right" questions to research implies a benefit-cost (efficiency) analysis for the prospective project that is essentially the same as for any other resource-using project, such as in water resources or manpower training. Thus, upon careful scrutiny the imaginative effort by the organizers of this session to break a monstrous evaluation task into two distinct evaluations fails to pass the test of separability.

Despite my conclusion regarding the simultaneity of judgments on which are the right (best) questions to ask and on the costs and quality of expected answers, I shall proceed. In the remainder of this paper, I try to identify the nature of the research questions

*University of Wisconsin-Madison. My thanks go to W. Lee Hansen, Michael Oineck, and Mark Schlesinger for their comments on an earlier draft. This research was supported in part by funds granted to the Institute for Research on Poverty at the University of Wisconsin-Madison by the Department of Health, Education, and Welfare pursuant to the provisions of the Economic Opportunity Act of 1964.

that have been posed in the economic education literature, and the nature of the questions that have *not* been posed. I will comment on whether the overall research program—the set of questions being asked—is what it “should” be, and attempt to point to researchable themes that are likely to have relatively high returns for research in this field.

I. Economic Education and General Education

One basic question is, why study *economic* education at all? What reasons are there to believe that the subject matter of economics is sufficiently special so that the voluminous general literature on teaching and education is not applicable to economics?¹ I have not seen this question posed in the more than 150 papers I have surveyed in the *Journal of Economic Education*, (*JEE*) and in the annual American Economic Association (*AEA*) sessions on economic education.² (There are, of course, papers on economics education published elsewhere, but my survey does not extend much beyond these two “official” *AEA* sources.) My point is simple. Is there not a substantial probability—indeed, perhaps not a presumption—that researchers studying economics education are “rediscovering the wheel,” posing and answering questions that have been answered previously in the more general research on education? For example, is it not likely that the effect on “learning” of, say, class size, or of the use of teaching assistants rather than more experienced professors, or of individually self-paced approaches rather than a traditional uniform, instructor-paced approach, is similar for all subjects? I do *not* assert that the answer is obvious and affirmative. I only question whether it is a “high priority” research

matter to devote substantial resources to general questions of teaching techniques; questions that are not specific to the teaching and learning of economics and that have been studied extensively in other subject matter contexts. It may well be true that, as one economist at an *AEA* session on economic education recently put it, “Educational production functions are at least as interesting as those for hybrid corn” (see Elisabeth Allison, p. 228). Nonetheless, it would not follow that production functions for *economic* education are efficient topics for economists’ research.

II. The Production Function for Economic Education

What research has been undertaken in economics education? Most of it is devoted to exploring some portion of the production function for economics education. Of the 159 papers surveyed, I count 102—essentially two-thirds—dealing either with how to define and measure outputs (23 papers), or with the effect on output of various alternative inputs (79 papers). This production function orientation is consistent, however, with the *JEE*’s goal as stated inside the front cover: “To promote the teaching and learning of economics in colleges, junior colleges and high schools by sharing knowledge of economic education.”

Table 1 presents the 79 input-output oriented papers according to the principal type of input the productivity of which was being studied, and according to the level of schooling. Since each paper was counted only once, while some papers touched on more than one input or school level, the table is an incomplete portrayal of the research foci.

I have classified the independent variables in the production function as capital, labor, students themselves, course content, and instructional methods (ways of combining inputs). An impressive variety of variables have been researched. I cannot judge whether some inputs that have received little or no attention are “worth” studying—for example, the output effects of the time of day that the class is held (but see Rolf Mirus), the color of

¹This is not to say that there is nothing special about the teaching of economics. Economists typically believe, for example, that people have more misinformation and biases concerning economics than about other subject matters. (Mark Schlesinger pointed this out to me.) Even if this is true (see Kenneth Boulding) the question would remain whether resources devoted to teaching economics should be deployed differently than in other subjects.

²For an excellent survey of research on educational production functions, see Eric Hanushek.

TABLE 1—NUMBER OF ARTICLES ON VARIOUS PRODUCTION FUNCTION RELATIONSHIPS FOR ECONOMICS EDUCATION, BY TYPE OF INPUT AND LEVEL OF SCHOOL

Type of Input	Elementary And High School	Junior College	College	Graduate School	Nonschool	All Levels
Capital						
Textbooks	2		2			4
Computers			9			9
Television, slides, etc.		1	5		1	7
						20 (25%)
Labor						
Instructors	6	1	4			11
Graduate Assistants			5			5
Consultants	1					1
						17 (22%)
Students (ability, motivation, family background, other students)	1		4			5 (6%)
Course Content (subject matter)	2	2	7			11 (14%)
Instructional Methods						
(ways of combining inputs)						
Games and Simulations		1	7			8
Learning contracts, self-paced instruction and programmed learning			11			11
Lectures			1			1
Course evaluations			4			4
Length of course			1			1
Class size			1			1
						26 (33%)
Total	12 (15%)	5 (6%)	61 (77%)	0	1 (1%)	79 (100%)

the classroom walls, or the seating arrangements.

A. Interaction Effects

What is probably a more serious omission is the lack of examination of interaction effects among input variables. It seems likely, for example, that a particular type of textbook (input *IA*) when used by graduate teaching assistants (*IIB*) will be more effective for low-ability students (*III*) than they would be for high-ability students. Similarly, games and simulations (input *VA*) may be differentially effective depending on whether instructors (*IIA*) or graduate assistants (*IIB*) are used and depending on the student's initial level of motivation (*III*).

B. Limited Scope

Another striking aspect of Table 1 is the overwhelming emphasis on teaching at the

college level (77 percent of the papers). The *JEE* goal, stated above, refers to "colleges, junior colleges and high schools." The scant attention of economics education researchers to high schools and junior colleges is noteworthy, given that half of young people do not go beyond high school, and that those who do go further are increasingly likely to go to a junior college. (Examples of research on economic education in junior colleges are Dennis Weidenaar and Joe Dodson, and Darrell Lewis, Donald Wentworth, and Charles Orvis. For a precollege focus, see Rendigs Fels (1977) and Thomas Duff.) It may or may not be true that the production function findings for the *college* population apply also to the junior colleges and high schools; the issue merits attention. Students' ability and motivation levels (as well as the variances in those levels) vary across the schooling levels; thus, the interactions of these students char-

acteristics with other, conventional inputs will produce, I hypothesize, different output effects depending on the level of school.

The narrow scope of teaching settings on which research has been published is also evident from the dearth of attention to the production function for teaching economics either in graduate schools (see, however, W. Lee Hansen and Robert Decker for models predicting success in graduate economic studies) or in nonschool settings such as in the home via television (see John Coleman) or via popular journalism (magazines and newspapers). How "effective," for example, are the syndicated newspaper columns of writers such as Sylvia Porter, the *Newsweek* columns by Milton Friedman and Paul Samuelson, the articles in magazines such as *Challenge* or *Public Interest*, or in daily newspapers? How effective—and for whom—are the efforts of private firms to provide "economic education" via newspaper advertisements (for example, Mobil Oil on energy issues)? These are unanswered—indeed, unasked—questions. Yet, the vast majority of people have not taken and never will take a formal economics course in any school, and they will be exposed to economics only through such informal media. Thus, the production function for learning economics outside traditional schools seems to warrant substantial exploration—assuming, of course, that economics is worth the opportunity cost of learning it. The omission of nonschool teaching and learning of economics from the *JEE* statement of policy is unfortunate.

C. Distributional Effects

I turn next to a related aspect of the production function work: the *distributional* effects of alternative course contents, input combinations, and instructional materials. These have been studied to some extent (for example, Richard Attiyeh and Keith Lumsden; Hansen, Allen Kelley and the author; Fred Thompson); yet, based on the evidence from the general literature on education that a given approach is likely to have substantially different effects on different "types" of

students, this dimension seems to deserve more scrutiny. Whatever the mean differential may be between the output effects of different inputs, examination of the variance about the mean may disclose systematic differences among students according to characteristics that are discernible at the outset of a course.

III. Outputs

A. Goals of Economic Education

The body of research presented in Table 1 focuses on the productivity of various *inputs*; the dependent variable—output—is generally taken as given, typically in the form of some test score. There is, however substantial other literature—not in the input-output framework—discussing the normative question of how *output* ought to be defined and measured. There are papers that discuss the "usefulness" of a specific output measure, particularly the Test of Understanding in College Economics (*TUCE*) (for example, see Darrell Lewis and Tor Dahl; Fels, 1977). Other concepts and measures of outputs on which papers have been published include changes in student political attitudes (see James Scott and Mitchell Rothman) the students' own judgment of effectiveness (see Kelley); learning "radical" economics (see Richard Edwards and Arthur MacEwen; John Gurley); and developing problem-solving abilities (see Fels, 1973). In addition, the durability or permanence of the effects, as distinguished from measures of effectiveness obtained at completion of the course, has received some attention (see Phillip Saunders; Saunders and G. L. Bach).

Overall, however, the question of what economics education ought to be aiming at—that is, which outputs should be produced—is a question that has received little rigorous analysis. The question of what kind or kinds of "economic education" to produce is a difficult one. Should it be ideologically oriented? Should it provide whatever "buyers" want? Who are the buyers—parents? taxpayers? students? Our custom-

any consumer sovereignty model appears to be of limited guidance here, given widespread consumer ignorance of the importance of economic knowledge, and given the external benefits from having a population that is more sophisticated in its understanding of economic processes. In economics education, as in many other "professional" markets, buyers are poorly informed regarding product quality. Even if buyers know their objectives, they may know little about the effectiveness of particular activities in achieving those objectives.

My references to "consumer ignorance" and to "external benefits," however, are scarcely more than assertions. I have seen little research that rigorously confronts the question of whether there is a market failure in the economic education market, with too few people studying too little economics or studying the "wrong" economics. The published research either asserts that more economics is good—and presumably is better than some unspecified alternative uses of student time and other resources—or else the research asks the narrower production function question of how effective one type of input is compared to another, without asking whether the output is worth producing. In volume 1 of the *JEE* George Stigler (p. 78) did pose the question, "Why should people be economically literate, rather than musically literate, or historically literate, or chemically literate?" I will resist the temptation to discuss his answer—except to note that musicians, historians, and chemists may see things differently.

B. Effectiveness vs. Allocative Efficiency

The domination of a production function emphasis in economic education research has obscured the related issue of the allocative efficiency of alternative input combinations. Many papers have examined the *effectiveness* (productivity) of various inputs, but rarely have the relative *costs* of the inputs been juxtaposed to the relative effectiveness, nor have the measures of effectiveness been translated into *values* of benefits. These questions

have seemingly been overlooked or, at least, slighted.

I find it surprising that among the (admittedly small number of) papers confronting the question of how to define the output or outputs of economic education, there has been so little attention to labor market effects in general, and earnings effects in particular. The contrast between the economic education literature and economics of education literature is dramatic. The latter has concentrated, typically within a human capital theoretic framework, on the relationship between education (meaning schooling) and earnings, virtually disregarding the *process* through which educational inputs produce the outputs that have value in the labor market. Another way of saying this is that the economics of education literature has viewed *earnings* as the value of outputs. Meanwhile, the economic education literature has concentrated heavily on the process of converting inputs into outputs in nonpecuniary forms, virtually disregarding the valuation of outputs.

One might have predicted a priori that the economic education literature would have included numerous efforts to assess the labor market value of economics training either directly or indirectly through its effect on, say, the probability of admission to law school. Why the economic education literature and the economics of education literature have been so divergent, and whether either, or both, or neither has pursued an "optimal" path are questions which I raise here, but will not pursue far.

C. Lifetime Effects

The human capital framework, within which much of the economics of education literature has been cast, has focused research attention on the investment aspect of schooling. The investment emphasis implies a *lifetime* perspective on the outputs of schooling. By sharp contrast, the economic education literature has concentrated overwhelmingly on the *immediate* outputs, those measured at the completion of the course. As pointed out

above, there have been a few noteworthy exceptions in which the durability of outputs has been considered, though even these have involved a horizon of only a few years or so (see Saunders; Saunders and Bach). It may well be exceedingly difficult to measure lifetime effects of exposure to economics, and this may explain the lack of attention to this question in the literature. (This would illustrate the interrelatedness of the "do-ability" of research and the formulation of research questions.) But the fact remains that little effort has been devoted to the measurement of lifetime effects.

IV. Incentive Structures

Another underresearched area is the nature of incentive structures facing teachers and administrators. Assume that 1) the production function research disclosed that certain inputs are more effective than others, 2) consensus was reached on appropriate measures of outputs (i.e., effectiveness), and 3) outputs and inputs were valued and showed positive net benefits from a change in current teaching practices. Would the changes occur? Are there incentives sufficient to encourage changes that are efficient (granting that such changes can be identified with reasonable confidence)?

These questions, it might be argued, transcend economic education. It would seem, however, that the responsiveness of teachers and administrators of economic education programs may or may not be the same as for those in noneconomics areas; at least this hypothesis cannot be ruled out, any more than can the hypothesis that variation in class size, or in the effectiveness of teaching assistants, or the use of television instruction differs as between economics and other subject areas.

The nature of incentives confronting teachers—of economics or of anything else, and at various levels of schooling—has received scant attention. There are possible incentives for instructors (a) to learn which changes are efficient, and (b) to make those changes. (On the latter point, studies of salary determination at universities, see, for example, John

Siegfried and Kenneth White. James Koch and John Chizmar have shed some light on the financial returns to scholarly research, teaching, and other uses of faculty time.) It is arguable that little is to be gained from research on *how* to "improve" teaching if the *incentives* to adopt improved methods are weak. It is also arguable, on the other hand, that incentives are weaker than they might be because there is so little agreement as to what constitutes efficient teaching; this, after all, involves the specification of goals in operational terms and the adoption of value weights for the multiple goals that surely exist. Thus, understanding *goals and weights* is one part of the research agenda for efficient innovation in education.

In any analysis of incentives in education the relationship between private costs and social costs (or returns) is likely to be crucial. As an illustration, consider the case of an economics instructor who is free (although many are not) to select any undergraduate textbook, and that a new textbook appears on the market. There may well be little incentive (financial, professional, or any other) to read the new textbook carefully enough to determine whether it is superior to the one being used; this, however, is not my principal point. What if the instructor knew—costlessly and with certainty—that the new book was "more effective" for all of his students; what would be the private and social costs and benefits of adopting the new book? Of course, more effective need not imply "more efficient."

From the students' viewpoint, the new book would presumably be preferred if it were more effective. Such a preference, in turn, embodies two deeper assumptions: (a) the similarity of student goals and of faculty goals for students, and (b) the absence of higher costs (time, effort, money) for using the new book that offset the benefits of increased learning.

Note, however, that while the student must incur the cost of reading whatever textbook is chosen—an essentially fixed cost—the faculty person bears an increased real social cost of changing, since he or she has lecture notes keyed to a textbook that has already been

read. With the benefits of change accruing to students while the costs are borne by faculty, the likelihood of market failure is substantial.

The market failure would disappear, however, if the instructor internalized the students' benefits. This might appear to be the case if the instructor acted as an idealized "professional"—that is, acted as the consumer's agent for maximizing the consumer's (student's) utility. Education is an example of a commodity—like medical care and legal representation—in which consumers are aware of their inability to judge quality, and so they place trust in the professional to act in their best interest. Even if the instructor were to behave, however, so as to maximize not his or her own utility but that of students (or parents, or taxpayers), it would not follow that efficient resource allocation would result. The reason is that the cost of switching textbooks (or, in general, of changing anything in the teaching process) is a real cost; if it were to be disregarded—as would be the case if the instructor were to act so as to maximize the consumer's utility—the result would be excessive change.

The market failure would also disappear if the reward structure were such that the instructor's pay were an appropriate function of the "value-added." Then, if students learned more from the new text, the instructor—acting in self-interest—would weigh the costs of changing books against the benefit, and would choose accordingly. Ideally, the rewards would be commensurate with the student benefits, and so—assuming away real external effects and other market imperfections—the instructor would be confronted with the real costs and benefits of change. The problems of developing such a reward system are doubtless great. It does not follow, however, that they are not researchable.

These remarks have been abstract; meat must be put on the analytic bones. I hope that the next time the economic education literature is surveyed there will be found more papers exploring incentives for innovation and efficiency—in both positive and normative dimensions.

V. Concluding Remarks

It is all too simple to find questions that one would like *other* researchers to tackle, as I have done here. Thus, I should close by reiterating my claim made at the outset that the selection of optimal research questions is, in principle, a matter of weighing benefits and costs, of comparing the value of having answers to the costs of obtaining them. If the costs are sufficiently high, it would be inefficient to research questions that seem important. Some of the questions to which I have pointed probably fail such a benefit-cost efficiency test, and so have received, quite wisely, little research attention; other questions, however, may pass it—at least for some researchers—and so merit more study. Once more we can conclude that "more research is needed."

REFERENCES

- E. Allison, "Three Years of Self-Paced Teaching in Introductory Economics at Harvard," *Amer. Econ. Rev. Proc.*, May 1976, 66, 222-28.
- R. Attiyeh and K. Lumsden, "Modern Myths in Teaching Economics: U.K. Experience," *Amer. Econ. Rev. Proc.*, May 1972, 62, 429-433.
- K. Boulding, "Some Observations on the Learning of Economics," *Amer. Econ. Rev. Proc.*, May 1975, 65, 428-30.
- J. Coleman, "Economic Literacy: What Role for Television?," *Amer. Econ. Rev. Proc.*, May 1963, 53, 645-52.
- R. L. Decker, "Success and Attrition Characteristics in Graduate Studies," *J. Econ. Educ.*, Spring 1973, 4, 130-37.
- T. B. Duff, "Basic Economic Concepts in the High School Curriculum," *J. Econ. Educ.*, Fall 1971, 3, 5-10.
- R. C. Edwards and A. MacEwen, "A Radical Approach to Economics," *Amer. Econ. Rev. Proc.*, May 1970, 60, 352-63.
- R. Fels, "A New Test of Understanding in College Economics," *Amer. Econ. Rev. Proc.*, May 1967, 57, 660-66.

- _____, "Developing Independent Problem-Solving Ability in Elementary Economics," *Amer. Econ. Rev. Proc.*, May 1974, 64, 403-07.
- _____, "What Economics is Most Important to Teach: The Hansen Committee Report," *Amer. Econ. Rev. Proc.*, Feb. 1977, 67, 101-04.
- J. Gurley, "The Principles Course: What Should be in it and Where Should it be Going?," *Amer. Econ. Rev. Proc.*, May 1975, 65, 431-33.
- W. L. Hansen, "Prediction of Graduate Performance in Economics," *J. Econ. Educ.*, Fall 1971, 3, 49-53.
- _____, A. C. Kelley and B. A. Weisbrod, "Economic Efficiency and the Distribution of Benefits from College Instructions," *Amer. Econ. Rev. Proc.*, May 1970, 60, 364-69.
- E. Hanushek, "A Reader's Guide to Educational Production Functions," work. paper no. 798, Instit. Soc. Pol. Stud., Yale Univ. 1977.
- A. C. Kelley, "Uses and Abuses of Course Evaluations as Measures of Educational Output," *J. Econ. Educ.*, Fall 1972, 4, 13-18.
- J. Koch and J. Chizmar, "The Influence of Teaching and Other Factors upon Absolute Salaries and Salary Increments at Illinois State University," *J. Econ. Educ.*, Fall 1973, 5, 27-34.
- D. R. Lewis and T. Dahl, "The Test of Understanding in College Economics and its Construct Validity," *J. Econ. Educ.*, Spring 1971, 2, 155-66.
- _____, D. R. Westworth, and C. C. Orvis, "Economics in the Junior Colleges: Terminal or Transfer?," *J. Econ. Educ.*, Spring 1973, 4, 100-10.
- R. Mirus, "Some Implications of Student Evaluation of Teachers," *J. Econ. Educ.*, Fall 1973, 5, 35-37.
- P. Saunders, "Does High School Economics have a Lasting Impact?," *J. Econ. Educ.*, Fall 1970, 2, 39-55.
- _____, and G. L. Bach, "The Lasting Effects of an Introductory Course: An Exploratory Study," *J. Econ. Educ.*, Spring 1970, 1, 143-49.
- J. H. Scott Jr., and M. P. Rothman, "The Effect of an Introductory Economics Course on Student Political Attitudes," *J. Econ. Educ.*, Spring 1975, 6, 107-12.
- J. Siegfried, and K. White, "Financial Rewards to Teaching and Research," *Amer. Econ. Rev. Proc.*, May 1973, 63, 309-15.
- G. Stigler, "The Case, if any, for Economic Education," *J. Econ. Educ.*, Spring 1970, 1, 77-84.
- F. A. Thompson, "Problems and Prospects of Economics Education in Community Junior Colleges," *J. Econ. Educ.*, Fall 1970, 2, 31-38.
- D. J. Weidenaar and J. A. Dodson Jr., "The Effectiveness of Economics Instruction in Two-Year Colleges," *J. Econ. Educ.*, Fall 1972, 4, 5-12.

Where is the Economics in Economic Education?

Richard B. McKenzie

Economists can be distinguished from other social scientists not so much by the topics they investigate as by the method (that is, the suppositions and theoretical tools) they employ in their investigations. Support for this position can be traced through economic literature, although it will suffice to note that Keynes defined economics as "a method rather than a doctrine, an apparatus of the mind, a technique of thinking, which helps its possessor to draw correct conclusions." Although the precise scope of this method is debatable, most economists agree that the choice calculus of the individual, as opposed to the group, is at the heart of the economist's view of the world.

Because the learning process for both student and teacher involves choices which are basically similar to those which they confront in their other pursuits, the tools of the economist have a clear and direct application to research in economic education. To date, however, economic education has been predominantly studied with the statistical tools and learning theories of professional educators. Given an almost total neglect of the conventional, economic approach in the subfield of economic education, one cannot avoid asking: Where is the economics in economic education? Do economists not have something to contribute, as economists, to the study of how they teach and what they teach?

In this paper I attempt to examine what economic educators are and are not doing. The focus is not on the education process, *per se*, but rather on how economic educators have studied the process and how I believe they *should* study it—as economists. In this context, the paper is openly critical of the status quo, more for what has been neglected than for what has been done, and it attempts to be persuasive to the extent that suggestions for redirecting economic education research are implied. Admittedly, the nature of the task I have set for myself requires that I draw generalizations concerning the work of an economic education profession which has many diverse elements; and space limitations dictate that my generalizations be more general and less embellished with qualifications than I would like. But I proceed on the assumption that we all share common interests and seek answers to many of the same questions. In the final analysis, my comments are valid to the extent that we share the same point of reference, and for this reason I begin within a preliminary statement of what I perceive that point of reference to be. If we disagree on the reference point, we will surely be at odds over what we as economic educators should do.

Richard B. McKenzie is Associate Professor of Economics, Clemson University. The paper was written when the author was at the Center for the Study of Public Choice, Virginia Polytechnic Institute and State University. He is indebted to his colleagues there, Robert Staaf, Nicolas Tideman and James Buchanan, for helpful comments and suggestions.

The Economic Approach and a Critique of Economic Education

If economics has any claim to being a science, it is to the extent that it has a positive as opposed to a normative tradition. The central concern of the economist has traditionally been with what is, and not with what ought to be. The individual, so long the cornerstone of economic theory, has been presumed to choose in accordance with his own preferences; no obligatory values or behavior patterns have been imposed upon him or presupposed for him. He is a free agent, operating within physical, genetic and social constraints. Economic theory in this sense is nonnormative, general and applicable where freedom of choice exists. It has explanatory and predictive power precisely because of, and not in spite of, assumed individuality. And it is virtually axiomatic within the profession that to know the individual is to understand the mechanism through which social phenomena (including education) occur; to bypass consideration of the individual choice calculus runs the risks inherent within the hit-or-miss, *ad hoc* theorizing so common in other disciplines.¹

Economic education, *per se*, has yet to be brought squarely within the positive tradition. Granted, many economists have made empirical studies of the education process as it pertains to economics; and with a reasonably broad conception of the discipline, such studies may be deemed economic in nature. Obviously, a major concern has been to improve the efficiency of the learning experience, which many think constitutes an economic problem, one which has been seriously ignored in the past. However, the theoretical models and assumptions undergirding many of these studies raise issues which must be carefully examined and which are far more complex than one may suppose. Let me briefly outline the nature of those issues in the hope that others will consider them.

The Problem of the Underlying Model

It seems fair to say that most studies conducted in economic education have been based on what may be called the "educational model" of student and faculty behavior. This model, which is used extensively in education, psychology and sociology presumes that student and faculty behavior is *determined* by genetics and the many environmental forces (social and physical) which are present in the educational setting and that these forces can be so arranged that student and faculty behavior is "shaped" to achieve the desired responses. Using this model, economic education has been treated as a mechanical process or as an input-output problem approximating that of molding a vase from clay. The individual (the student), who has in a sense been assigned the passive role of the clay, has been devoid of any hint of the rationality normally associated with *homo economicus* and has, perhaps, been lost behind the computer and in the discussion of the performance of the "group" or "class." Without explicitly providing for choice, and by viewing student and faculty behavior as that which is revealed in data collected after the learning experience has taken place, the assumption implicitly and often made is that the independent variables entered on the right-hand side of regression equations work directly on the dependent variable on the left-hand side—that the individual students *respond*, although imperfectly, to the punching, pulling and molding of the master craftsman (the instructor). Such a casting of roles, no doubt, elevates the importance of the economic educator.

Use of the education model of student and faculty behavior needs to be seriously questioned for several reasons. First, it provides educational researchers with a "mind-set" which views the learning experience and its outcome, not as that which emerges from the interaction of independently motivated individuals with the capacity to act upon (as opposed to respond to) social forces, but rather as an experience which must be controlled to obtain an objective which is largely derived externally from the learning experience and is imposed on it. That which is defined as "good" or "efficient" is not necessarily that which is defined to be "good" or "efficient" by the participants in the learning experience. Contrast this perspective with the economist's vision of the market in which the values of the individual market

participants are the criteria for determining the efficiency of the market process. When two people trade their wares, the exchange is "efficient" in the sense that a welfare improvement has occurred because the market participants define the exchange to be *good*. They enter into the exchange and benefit from it, not because the market environment has been molded so that they will enter into it, but because they *perceive* the exchange to be beneficial.

The economist's view of the emerging market process is an argument against control, not for it. On the other hand, the perspective of the education model, with its externally derived efficiency criteria is an argument for control and, to some extent, a denial of the values of the individual participants. To the extent that any mind-set facilitates thinking along certain lines, it reduces the cost of drawing certain types of conclusions and increases the frequency with which those conclusions are deduced. The conclusion so easily drawn, using the perspective of the education model, is that we need to specify *our* goals more carefully and determine with greater precision *our* standards of performance. Using the education model, it seems likely that researchers will be more interested in seeking out "common goals" for the economic education process and in extending controls over that process than will researchers who are imbued with the conventional economic model in which the reference point is the value set of the individual educational participants and the "goodness" of the process is defined by the mutually beneficial trades which occur between participants.

The reader may be concerned at this point. I agree that there is also a problem in viewing the learning process completely as a market phenomenon. It is clear that young children required guidance from their parents. They learn what to do and what not to do, what is "good" and what is "bad," from parental values which implicitly determine the deliberately imposed constraints on their behavior. However, notice that the goals which are established in rearing a child at home are set at a highly decentralized level and that they vary markedly from household to household and even more markedly from neighborhood to neighborhood. The educational model may therefore be quite applicable to the problems of parents, and it can justifiably be extended to the elementary school and applied to the teaching of certain subjects which are acceptable to (practically) everyone. Unanimous agreement on collective issues, such as goals for education, is the public counterpart of mutually beneficial trades in a market setting; it insures that only Pareto efficient collective decisions are made. The general agreement on the goals of certain kinds of primary education, taught in certain ways, insures that the outcome is desirable. My concern is whether or not we can take the argument that is used for the application of the education model to home-produced education and to elements of primary education and apply it to higher education and subjects like economics over which there is broad disagreement as to what it is, what it should be, and what should be done with it. These extended applications of the education model involve very large and diverse groups of people, and the so-called "common goals" require that collective as opposed to individual decisions be made within the relevant group.

Using the education model as their frame of reference, economic educators are understandably interested in defining the goals of economic education and of specifying in very detailed ways how the attainment of those goals will be measured. My concern is whether or not *the* goals of *economic education* (as opposed to goals established at highly decentralized levels) and whether or not the means of measuring the degree to which those goals are achieved are legitimate problems when considered in the context of a *national* economic education movement. It seems to me that the mind-set which one carries with him in his educational research will be quite influential in determining how he will react to what I have just said. The mind-set of the education model will lead the researcher to the establishment of goals, since goals are necessary for the development of further educational research. The mind-set of the economic model, on the other hand, will lead the researcher to question any collectively established goals for economic education; economic education is not that which it *should* be, in

some collectively defined sense. Rather, it is that which emerges from the educational process, and that which emerges continuously evolves in many different ways. Indeed, academic freedom is a property right which we all enjoy and which recognizes the existence and need to promote diversity in the learning experience. This seems to be the spirit in which Jacob Viner defined economics as that which economists do.

Later, I will argue that, if one accepts the "citizenship argument for economic education," consistency mandates that one be concerned with common goals.² Having established common goals, the relevant research model is the education model. The economist's model of learning as an emerging process which is dependent upon individual values must be rejected. I will also argue that the citizenship argument is, from an economic perspective, a questionable basis for promulgating economic education. Again, these arguments relate back to the model that is used and attack the foundations of the research which most of us undertake. The irony of this line of analysis is self-evident. As economists we teach about the market process and use elaborate models to discuss that process; if we start our research with externally derived goals, however, we cannot use the models which we develop in class to evaluate that which we do in class.

Setting this line of argument aside for the moment, other problems with the education model can be noted. We have indicated that use of the education model presupposes some set of goals and a means of measuring achievement of these goals. However, as my colleague Robert Staaf has pointed out, goals and testing techniques cannot be expected to fall from heaven like manna [20]. Rather, goals and testing techniques must be established through some collective decision process. The goals which do emerge from the collective decision process will be dependent on the voting rule that is used in the process. If the goals established by collective decisions vary with the voting rule which is employed, then there is reason to question that the goals which are established, given the voting rule, are meaningful. Using the public choice perspective, Staaf goes on to question whether or not efforts to determine what is good teaching in economics based, for example, on a simple majority voting rule, are very meaningful. In questioning the collective basis for determining the voting rule, he also questions whether the educational model is very useful. Even if one rejects Staaf's conclusion, his paper serves the important purpose of suggesting that economic educators must not only be concerned with goals and test instruments, but also with the means by which those goals and tests are established, that is, the collective decision-making process.

If the goals for economic education are established before research in economic education begins, it is all too tempting for researchers to conclude that the establishment of benefits from economic education justifies the establishment of programs in economic education. This is, in essence, the mistake which Milton Spencer makes when in his introductory textbook he writes to the student: "But in any case, whatever career you pursue, a knowledge of economics will make you a far more effective citizen, and this alone justifies the time devoted to its study" [18, p. 5]. Using the education model as the point of reference, it is easy to understand why the costs and benefits, and particularly the costs, of economic education are seen so seldom in economic education literature and why estimates of their values are so rarely introduced as arguments in regression analysis. In class we are constantly concerned with questions of whether or not government regulation, antitrust actions, tariffs, and criminal justice systems are worth their cost. Is it not reasonable—if only for the sake of consistency—for us to ask the same type of questions about what we teach? Shouldn't the opportunity value of the student and instructor's time be considered in any decision about making economics a graduation requirement for high school or college or in any attempt to employ, for example, self-paced learning methods or criterion reference grading procedures or in evaluating how much students learn?

Several economists have begun to consider the choice calculus of individual students and faculty in determining educational outcomes. In this regard, Robert Staaf in an important article

in the *Journal of Economic Education* has led the way [19]. Whereas many economic educators had begun to wonder if anything will work in upgrading economic literacy, Staaf demonstrates with elementary tools of analysis that many of the techniques which have been used in the classroom on an experimental basis may in fact be working. However, by not considering student choice, we may have simply failed to make our studies sufficiently general and have failed to view the learning process as a problem in time allocation for the student. The student may transfer the efficiency benefits acquired in one course to his study of other subjects and to other noneducational activities. I suggested in a paper written subsequent to Staaf's that the professor may do the same thing; that is, he may transfer efficiency benefits to other courses, research and leisure activities [12]. The conclusion which may be drawn is that, even if a change in educational methods results in greater efficiency in the learning process, the student and faculty may disperse the efficiency benefits of any change over so many activities that it is impossible for present statistical techniques actually to show that efficiency benefits have been achieved. More recently, Allen Kelley [9] and Paul Kipps [10] have also made contributions in the application of economic models to economic education.

Finally, the education model can be criticized on the grounds that it really is not a model of behavior in the sense that it yields refutable hypotheses. At its base, it assumes that student learning is a function of the numerous forces which come to bear on the student in the classroom. There is, therefore, no *a priori* basis for concluding that any one of these forces is either less or more significant than any other, nor is there any *a priori* basis for determining the directional influence of any force. The directional influence of any variable is simply that which is established in the regression equations. Statistical procedures are used to develop the "model" as opposed to testing one that is sufficiently well-defined to yield refutable predictions. When an unlimited number of "hypotheses" can be deduced from any theoretical frame of reference, we must question whether or not any one "hypothesis" is anything more than conjecture. Jerome Katz reminds us: "A hypothesis that fails to conform to the known facts is simply false, but a statement that fails to assert something beyond this is not a hypothesis at all but merely a report of past experience." It seems to me that in economic education we have predominantly been reporting our past experiences.

The "Citizenship Argument" for Economic Education

One of the most widely used arguments employed to encourage public and charitable support of economic education is that the study of economics will contribute to "improved citizenship" or a "better citizenry." By this it is generally meant that the coursework will in some way raise students' awareness of the political, economic and social system of which they are a part, increase the intelligence with which they vote, change attitudes, and increase people's participation in the political process.

The citizenship argument has been articulated in a number of sources, but space limitations prevent one from showing the variety of ways the argument is used. [See, for example, 3 and 8.] I can only stress that in all its varied forms, the citizenship argument implies a strong presumption that economics will affect students' political-citizenship ability and behavior. In its elementary form the argument is correct in the sense that it is internally consistent: *If* students take economics, *if* they learn the subject matter and retain what they learn, and *if* they employ what they know in their political-citizenship behavior, the result can be an improvement—as defined by economists—in the voting decisions of the public and a possible increase in the efficiency of government. The general welfare, as the argument goes, can be increased.

The citizenship argument must be questioned on several counts, and I have undertaken that task in other papers [13-15]. Therefore, my comments will be brief here.³ The citizenship argument suggests that through transforming student-citizen behavior, economic education will produce a public good; and, indeed, if all act on the basis on the economic principles they

study, this can be the case. Herein lies the *economic* argument for *public* support of economic education. However, we must recognize that the argument is dependent upon *group* behavior and an assumption of *group* rationality. In class we teach that national defense and a host of other public goods must be produced publicly because individual citizens, who may recognize the benefits of the public goods, will have no incentive to provide them on a voluntary basis. A person may become a free rider because he knows that his individual contribution to the public good will have an insignificant effect on the benefits which he, himself, receives. Further, voluntary collective action to produce public goods is not likely to occur because of the extensive transaction costs involved in trying to bring about collective behavior from a large number of people who have an incentive to free-ride on what others may do.

Does this argument also apply to the public good of economic education? The public choice theory developed primarily by James Buchanan [2], Anthony Downs [5], and Gordon Tullock [21] suggests that it does. They argue that because the individual citizen has essentially no effect on the outcome of the political process—that is, the type of public policies that he wants—he has no incentive to incur the cost of becoming politically intelligent and using the intelligence which he has acquired in his political behavior. Unfortunately “intelligent” political decisions by the electorate constitute a very special kind of public good: If those decisions are to be free, then they must be made by individuals acting independently of one another. They cannot be produced like national defense; rather each individual must be free to determine voluntarily what his contribution to the public good will be.

Through forced instruction and artificial incentives like grades given for classroom work, we may be able to overcome the free rider tendency to learn the type of economics designed to improve citizenship. However, we must recognize that there is a wide gulf between the classroom experience and the voting booth. Before the public benefits of economic education can be acquired through the political process, the student, upon leaving class, must have sufficient incentive to maintain the human capital stock that he has acquired in class. Furthermore, after he has left the classroom he must have sufficient incentive to employ that which he has learned in determining the political positions of candidates and the possible consequences of the alternative government policies that are proposed. Given the fact that most of us teach methods, as opposed to settled conclusions, the cost of retaining and using economic education in the political arena can be quite significant. Public choice theory predicts that the typical individual student-citizens will not have sufficient incentive to incur these costs. At the very least, public choice theory offers economic educators an hypothesis to refute and challenges them to refute it. The evidence which is based on a study I undertook during the 1974 election supports the public choice hypothesis [13]. Support for the public choice hypothesis, which actually applies to a number of courses of study, is also found in the political science literature [6 and 11].

Finally, economic education must overcome the tendency of people, in spite of what they know about the economic merits of legislation, to vote their own private interests. I doubt very seriously that there are many textile executives trained in economics who will vote against tariffs on imported textiles. Indeed, the people who are most likely to remember their economics instruction are those who have an interest in manipulating government policy to serve their own ends. I suggest we look more carefully into how economic education is used by political entrepreneurs.

Regardless, the citizenship argument poses a real dilemma for economic education. If the ultimate objective of economic education is to produce a “better citizenry,” then we must ask what a better citizenry is and how the better citizenry can be produced. This requires that we seek common goals and procedures. If we do not address such questions and allow the economics profession to proceed in the development of the discipline in many directions, none of which is defined to be superior to the others, then we effectively do not have a reference point

from which to judge whether or not economic education has contributed anything to "improved citizenship." There is no means for establishing that a public good, as opposed to a public bad, has been produced; there is no way to establish whether or not the benefits of the public good that is produced are worth the cost. This is because in an unconstrained environment, one in which no common objective has been established, any and everything counts. However, if we agree that common goals must be established, then we immediately confront the problems of collective decision-making within the economic education system. It is not clear that the goals decided by collective decision are necessarily superior to individually established goals which are at variance with the collectively established goals and procedures. Collectively established goals may reflect the views of the median voter group within the profession, but that does not necessarily make these goals superior to the goals of others. Again, if all economists were in perfect agreement as to what economic education should do, and if economists did not have private interests in establishing goals, the problems of collective decision-making would not arise. However, I question whether we are as similar in our outlooks and as disinterested as many seem to suppose.

Summary

Individual economic educators have made some interesting and useful changes in their classrooms. However, the purpose of this paper has been to raise issues which have largely been avoided in the past and cannot be avoided in the future. I believe that we must critically reexamine the presuppositions of much of our work and make use of research being done by economists outside the economic education profession *per se*. Economic educators should make greater use of the tools of analysis which they teach in the study of what they teach. To do this, they must step back, away from the classroom and the particulars of their teaching experiences, and consider the student and the teacher as rational individuals who look upon economic education as one of many possible goods over which they must allocate their resources to maximize their *individual* welfares.

We must ask some tough questions. Do voters really have sufficient incentive to become and remain economically literate and to employ what they know in the political process? Economics is one of a number of subjects which have potential external benefits. Assuming that students cannot learn everything, what claim does economic education have on public resources relative to other subjects? Can we argue for the introduction of economic education without opening up the public school curriculum to exploitation by all interest groups who believe that their subjects rightfully have greater claim to resources than economics? I note that the elementary school curriculum has been opened up to special educational interest groups, and the curriculum has literally become a hodgepodge of subjects. This development is taking its toll on basic educational skills. Finally, we must ask: Is the cost of raising the public's economic literacy, as defined, for example, by the TUCE, greater than, equal to, or less than the benefits achieved in government policy because of economics instruction? Although we may wish (or believe) that the benefits exceed the costs, this indeed may not be the case, particularly in the light of the problems the profession has encountered in raising the understanding of students. These are positive questions which go to the heart of what many perceive to be the justification for our existence. Answers may not only tell us what we are doing wrong but may also indicate why we have not done more.

Footnotes

¹All that I have said so far is not meant to suggest that there are no normative elements associated with economic theory. Coats [4], Nabers [16], Rothenberg [17], and Heyne [7] have all alerted us to the implicit values embedded in the economic method, and Boulding [1] has reminded us that the acceptance of the economic method as a mode of inquiry involves the acceptance of what he calls the "economic ethic," that is, the principle that the "goodness" or "badness" of any social event should be founded upon an evaluation of its rewards and costs as perceived by the relevant individuals.

²Succinctly stated, the "citizenship arguments" presupposes that people need economic education to make them more intelligent voters and, therefore, better citizens.

³In these other papers, I have dealt [13] with counter-arguments to the public choice theory about to be briefly described and [14] with the design of courses for rational students.

REFERENCES

1. Kenneth E. Boulding, *Economics as a Science*. New York: McGraw-Hill Book Company, Inc., 1970, pp. 117-38.
2. James M. Buchanan, *Fiscal Theory and Political Democracy*. Chapel Hill: University of North Carolina Press, 1960.
3. *Centers for Economic Education: A Handbook*. New York: Joint Council on Economic Education (no date).
4. A. W. Coats, *The Role of Value Judgments in Economics*. Charlottesville, Va.: Thomas Jefferson Center for Studies in Political Economy, University of Virginia, 1964.
5. Anthony Downs, *An Economic Theory of Democracy*. New York: Harper & Row Publishers, Inc., 1957.
6. Daniel B. German and Marvin K. Hoffman, "North Carolina Youth and the Legal System," project report for the U.S. Department of Justice, the Law Enforcement Assistance Administration and the Western Piedmont Council of Governments, September 1973.
7. Paul T. Heyne, "On the Positive-Normative Distinction." Paper presented at the annual meeting of the Southern Economic Association, November 1974.
8. Robert V. Horton and Dennis J. Weidenaar, *The Goals of Economic Education: Adelphi-Like Inquiry*. Lafayette, Ind.: Institute for Research in the Behavioral, Economic, and Management Sciences, Purdue University, 1974.
9. Allen C. Kelley, "The Student as a Utility Maximizer," *Journal of Economic Education*, 6 (Spring 1975), 82-92.
10. Paul H. Kipps, "The Use of Course Evaluations to Influence Teaching and Research Activities," *Journal of Economic Education*, 6 (Spring 1975), 93-98.
11. Kenneth P. Langton and M. Kenneth Jennings, "Political Socialization and the High School Civics Curriculum," *American Political Science Review* (September 1968).
12. Richard B. McKenzie, "Faculty Choice, Educational Innovation, and the Public Good of Teaching," Boone, N.C.: Department of Economics, Appalachian State University, 1972.
13. ———, "Political Ignorance: An Empirical Assessment of Educational Remedies," *The Frontiers of Economics* (Summer 1977).
14. ———, "Politics, Learning, and Public Goods Education," *The Frontiers of Economics* (Summer 1976).
15. ——— and Robert J. Staaf, *An Economic Theory of Learning*. Blacksburg, Va.: University Publications, Inc., 1974, Chapter 6.
16. Lawrence Nabers, "The Positive and Genetic Approaches," *The Structure of Economic Science: Essays in Methodology*, Sherman Roy Krupp, Editor. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964, pp. 68-82.
17. Jerome Rothenberg, "Values and Value Theory in Economics," *The Structure of Economic Science: Essays in Methodology*, Sherman Roy Krupp, Editor. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1966, pp. 221-242.
18. Milton H. Spencer, *Contemporary Economics*, 2nd ed. New York: Worth Publishers, Inc., 1975, p. 5.

19. Robert J. Staaf, "Student Performance and Changes in Learning Technology in Required Courses," *Journal of Economic Education*, 3 (Spring 1972), 124-29.
20. ———, "What Is Good Teaching?" Paper presented at the annual meeting of the Southern Economic Association, November 1975.
21. Gordon Tullock, *Towards a Mathematics of Politics*. Ann Arbor: University of Michigan Press, 1967.

Student Performance and Changes in Learning Technology in Required Courses

Robert J. Staaf

Considerable controversy and confusion exists in this journal concerning the effectiveness of various input variables such as different teaching methods, textbooks and class size on student performance in introductory economics courses. In summary, the evidence suggests that these variables have an insignificant effect on student performance or that the available evidence is not conclusive. Possibly as a result of these inconclusive data and financial constraints, departments in many universities are adopting a policy of increased student/teacher ratios (large auditorium lectures) for "required introductory" courses.

This article develops a model which offers a new approach to analyzing student performance and may provide insight into measuring the impacts of a change in the input variables on the learning process. It is interesting to note that most of the changes or innovations have occurred in introductory courses which for many students are "distributive requirements." The analysis suggests that for many students these course requirements may be "inferior goods," thereby leading to unexpected behavior as a result of a change in the input variables.

Robert J. Staaf is Post-Doctoral Fellow and Visiting Lecturer, Center for Public Choice/Department of Economics, Virginia Polytechnic Institute and State University.

SOURCE: *Journal of Economic Education*, vol. 3, no. 2, Spring 1972, pp. 124-129. Reprinted with permission of the author.

Consider a hypothetical student, Albert, who is assumed to have average or moderate aptitudes in all fields of study. Or alternatively, Albert has no comparative advantage or disadvantage in any particular discipline. Assume Albert is a full-time student, carrying some minimum number of semester hours (e.g., 4 courses or 12 semester hours).¹ Albert's moderate or average aptitudes are illustrated by Figure 1.²

Assume the vertical axis represents knowledge in field (a) (e.g., social science) and the horizontal axis represents knowledge in field (b) (e.g., natural science). For expository reasons assume this knowledge can be measured and ranked objectively where a_4 represents greater knowledge than a_3 and so on.³ Suppose Albert has an equally divided course load between (a) and (b).⁴ Assume within Albert's time constraint (i.e., semester) and given his aptitudes and present achievements (a_0, b_0 at the origin) he is able to achieve any bundle of knowledge (combination of a and b) on the line ranging from x_1 to x_5 . Under these assumptions Albert's achievement at the end of the semester may range from a_4 and b_0 to b_4 and a_0 or any linear combination of (a) and (b) such as a_3, b_1 ; a_2, b_2 ; a_1, b_3 .⁵ These assumptions allow us to specify the attainable set and the trade-off possibilities between (a) and (b). A change in either Albert's aptitudes or time constraint increases or decreases the attainable set. Achievement on the boundary is guaranteed if we assume Albert aspires for (a) and/or (b).⁶ The particular bundle that Albert selects will be a function of his relative preferences for (a) and (b).⁷ Note that the bundle selected reflects Albert's allocation of his time (semester) between the fields (a) and (b).

Any bundle on the boundary of the attainable set may be thought of as being similar to Becker's full income approach [2, pp. 497-498]. Full achievement of any bundle (on the boundary) is attained by Albert devoting all his time and other resources at his command to learning activities with no regard for consumption or leisure activities. As with Becker's full income model, not all the student's time would usually be spent in studying and attending lectures. Time expenditures for sleep, eating and even some leisure are presumably required for efficiency in learning.

Now consider grades that reflect relative achievement levels in terms of (a) and (b). If we assume that a_4 or b_4 represents an "A"; a_3 or b_3 a "B"; a_2 or b_2 a "C"; a_1 or b_1 a "D" and a_0 or b_0 an "F" then the boundary of the attainable set represents a "C" average regardless how Albert allocates his time.⁸

This highly restrictive model reveals that a minimum G.P.A. of C can be attained by Albert allocating his time among a number of alternative achievement bundles (i.e.,

¹The analysis is not changed by assuming a part-time student.

²Aptitudes are defined as learning *rates*. Aptitudes are not assumed to involve a capacity or *stock* concept.

³Criterion reference testing may approximate this idea. Knowledge is assumed to be broken into "bits" of information for which one has knowledge or does not have knowledge. This criterion is void of norms (e.g., a class or national norm).

⁴This assumption is necessary since we are using a two-dimensional graph. The analysis can easily be extended to n dimensions.

⁵A constant ability or aptitude rate of substitution is only one possibility. See [1] for an analysis of decreasing and increasing rates. While the ability rate of substitution may not be constant, there appears to be evidence that suggests it is negative.

⁶Aspirations may be thought of as the dominance or nonsatiation axiom used in axiomatic choice models. Again, see [1] for a fuller account of this notion. As will be noted later by imposing grades over (a) and (b), the boundary is guaranteed by the assumptions of Albert's moderate aptitude and preference for college and a degree over not being in college and no degree.

⁷More specifically his personal rate of substitution. The allocation of his time and, therefore, his achievement in (a) and (b) is determined when his personal rate of substitution equals his ability rate of substitution.

⁸It is unlikely given grading practices based on a "normal curve" that the attainable set boundary is linear; rather it is more likely to be convex. Therefore, a student strategy of maximizing G.P.A. will not lead to devoting all his time to one field.

x_1, x_2, x_3, x_4, x_5). Absolute specialization in either field (x_1 or x_5) is unlikely under the conditions that Albert is required to take both (a) and (b) courses in a semester. A corner solution (all F's in one field) would mean Albert having to earn additional credits towards the degree. These additional credits would entail further time expenditures. Therefore, for practical purposes we can ignore x_1 and x_5 as being deliberately chosen. The above model assumes Albert is endowed with moderate or average aptitudes. If Albert were endowed with high aptitudes in all areas, he would be able to attain an A average *without* having to sacrifice A achievement in one field at the expense of A achievement in another field. The bundle x^* on Figure 1 represents the well-endowed student. A trade-off curve paralleling x_1-x_5 could be drawn that intersects x^* . However, x^* is a determinant solution if this student is a (G.P.A.) maximizer. Achievement beyond a_4 (e.g., $a_4 + \alpha$) comes at the expense of a lower achievement in b which is less than b_4 . The bundle $x^*(a_4, b_4)$ represents a straight A average while any other bundle on the boundary (dotted line) intersecting x^* would represent a lower G.P.A. (e.g., $a_4 + \alpha$ and $b_4 - \alpha$ does not yield an A average). The marginal product in terms of "grades" (not achievement) for the well-endowed student is negative for a reallocation of time that deviates from x^* . Between these two extreme cases of moderate aptitudes and exceptional aptitudes we might expect some trade-offs. In addition, even the exceptional students who have accumulated enough high grades in past semesters may to some extent trade potentially high grades for lesser grades in the current semester *and* more leisure and consumption time. Some students may have a high aptitude or above average (e.g., B) aptitude in one area and a below average (e.g., D) aptitude in another area. In the absence of course requirements, a student may be assumed not to choose any courses for which his aptitudes are low unless his preferences are strong for this field and he is willing to pay the price of a lower G.P.A.

For the moderately endowed student, a trading off of low grades (e.g., D's) in one area for high grades (B's) in another area may be quite rational. For example, courses (a) and (b) in Figure 1 may be interpreted as "major" course requirements and "distributive" course requirements, respectively. Since the choice of a degree bundle is in a large part a preference for a particular major, and distributive requirements are beyond

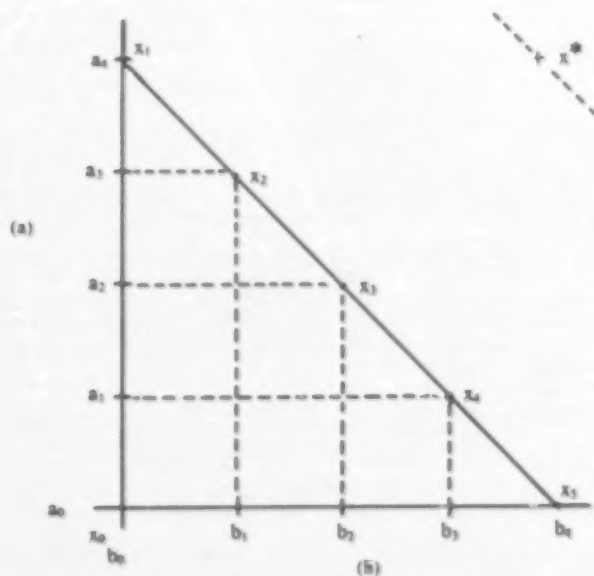


Figure 1

the control of the student, one would expect that the major course requirements are preferred over distributive course requirements. Similarly we can assume that graduate schools and most prospective employers would weigh performance in the major field over performance in the distributive courses.

David Riesman provides some evidence that tends to support a trade-off policy even for those students who are well endowed (e.g., x^*) [3, p. 42]. After trying in vain to persuade some of his students that they could think less about grades and more about education, Riesman says that one of his graduate students at Chicago finally did a thesis which documented his arguments. The student asked departments which graduates they had recommended for jobs, or advanced training or fellowships and then interviewed the students and looked at their grades.

"He concluded that those students frequently fared best who were not too obedient, who did not get an undiluted, uncomplicated, straight-A record. (The straight-A students, in fact, sometimes slipped away without anyone's noticing.)

The students who were most successful were a bit rebellious, a bit offbeat, though not entirely 'goof-offs'; these were the students likely to appeal to a faculty member who had not entirely repressed a rebelliousness of his own that had led him to be a teacher in the first place, a faculty member who was looking for signs of life, even if they gave him a bit of trouble at times. To be sure, such a student had to do well in something to earn this response, but he was often better off to have written a brilliant paper or two than to have divided his time, as an investment banker his money, among a variety of subjects."

The model suggests that students may not exert very much effort in some classes even in spite of the fact that the faculty has the power to give low grades which affect a student's chance of surviving in school. Furthermore, if low grades can be neutralized, the model suggests that students reveal their preference intensities over their course work. Therefore, it is perfectly rational for students to decide on different strategies in the allocation of their time. For the moderately endowed student a trade-off in favor of "major" course requirements would appear to be the most optimal allocation of his time. In addition, student behavior that in effect anticipates low grades in some courses does not necessarily mean that the student does not have aspirations for these courses, it simply means that given his preferences, aptitudes and time constraint he has chosen to behave this way to maximize his satisfaction. An increase in either aptitudes or time may change his behavior. According to the model developed, there are several reasons why students may be more motivated (expend more time) in major courses, whether they be required or electives. Of course, electives are by definition the prerogative of the student and therefore would presumably be chosen on the basis of his preferences. Therefore, student behavior in terms of time allocation may explain in part the evidence that tends to suggest that large auditorium classes (usually required courses) are no less effective than classes with smaller student/teacher ratios. If the student has decided on a particular allocation of time at the outset (beginning of the semester), it may be questionable whether the professor is able to affect even marginally his allocation choice given student preferences and constraints. If the introduction of new teaching techniques or texts is designed to simply change student "preferences" and, therefore, a student's time allocation, the model suggests that such techniques may not be very successful in required courses since trade-offs must be made in more "preferred" areas such as a student's major.

Technological Changes

On the other hand, assume that these techniques are effective in extending the boundary of the attainable set. That is, changes in teaching techniques, textbooks and class size really do make a difference given a student's aptitudes. Figure 2 is identical to

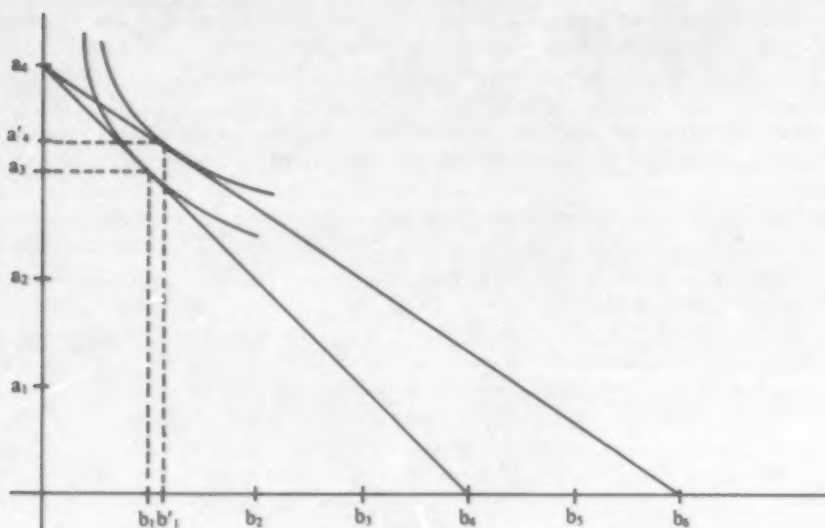


Figure 2

Figure 1 except that a change in teaching techniques or materials is introduced which is illustrated by the dotted line.

Assume (b) represents required "distributive" courses and (a) "major" courses. Techniques that are technologically effective are defined as increasing a student's "apparent" aptitude in the technologically affected area or course. These techniques are assumed to be external to the student and do not require increased student inputs such as time and intellectual effort. Our hypothetical student is now able to achieve b_6 if he devoted all his time to (b). If performance standards (correspondence between achievement and grades) do not change, the effect of introducing learning technology in area (b) is to lower the relative price of (b) represented by the dotted line. Indeed this may be the intent of introducing these techniques in required courses so as to induce a substitution effect towards (b), thereby tempting students to specialize (major) in (b).

However, an income effect is also associated with the relative price change in (b). The introduction of new techniques in (b) allows the student to allocate more time to (a), thereby increasing his grade in (a) without affecting his prior achievement level or grade in (b).⁹ Given the indifference map illustrated in Figure 2, the substitution effect is almost completely offset by the income effect.¹⁰ That is, courses in (b) are inferior goods. The net increase in achievement resulting from the change in technology of field (b) is $(b'_1 - b_1)$ which may not be statistically significant.¹¹ However, the technological change allows the student to allocate more time to his major field (a), thereby allowing him to increase his achievement $(a'_4 - a_3)$ and grades without significantly affecting what would have been his achievement level (b_1) and grades in the distributive courses (b). The assumption that required courses may be inferior goods for some students does not seem to be totally unrealistic. Therefore, studies that concentrate on changes in achievement levels in the technologically affected courses while ignoring effects in other courses may not find statistical differences if the courses are inferior goods.

⁹The solid line may be taken to represent a control group of students with similar aptitudes to those of the experimental group (dotted line) that is introduced to the new learning technology.

¹⁰Note that preferences are defined over a field of choice that represents achievement and grades.

¹¹In addition, it is conceivable that preference maps may be such that Giffen's Paradox occurs.

Some Suggestive Evidence

Studies on pass/fail tend to support trade-offs in a student's time allocation pattern over his course load. Note that pass/fail is not assumed to increase the student's attainable set in terms of achievement.¹² This evidence only suggests that students do make trade-offs when given the opportunity to do so at lower costs. A study at Dartmouth College by Feldmesser [4, p. 63] revealed that the most distinctive characteristic of the option was that it was a way of reducing the burden of distributive requirements.¹³ Users of the option tended to receive a full letter grade lower than nonusers regardless of a student's cumulative G.P.A. Further evidence suggests that of two students (of similar abilities) taking a course in their major field, the one using the option in another course would average about half a grade higher than the one not using the option while the grade deficit of the user in his option course does not affect his overall G.P.A. [4, p. 116]. While high G.P.A. students more or less made up in a course taken for the major what was lost in a course taken in the option, no such compensation occurred among low G.P.A. students. The lower achievement effects of the option in the course in which it was being used seemed to spill over into other courses such as their major [4, p. 133]. The time released seems to have been expended in other activities besides course work.

The above-mentioned evidence is highly suggestive of students revealing their preferences over their course work when a change in the relative prices is brought about by the introduction of a pass/fail system. The evidence supports the notion that distributive course requirements are inferior goods.¹⁴

Conclusions

The analysis suggests that the role of distributive course requirements requires a study in itself. Statistical studies that focus only on one course from a student's bundle of courses are likely to overlook the spillover impacts of a technological change in teaching techniques. It is not at all clear that these teaching techniques are efficient in the sense of increasing a student's attainable set or are simply attempts to change student preferences.¹⁵ In any event, the model presented suggests a new methodological approach to determining if these techniques are efficient by focusing on a student's time allocation pattern and substitution and income effects.

REFERENCES

1. Robert J. Staaf, *Student Choice: A Theoretical Analysis of Higher Education with Some Empirical Evidence Considered*. Unpublished doctoral dissertation, Temple University, 1971.
2. Gary S. Becker, "A Theory of the Allocation of Time," *Economic Journal*, 75 (September 1965), 493-517.
3. David Riesman, "Where is the College Generation Headed?" *Atlantic Monthly*, 207, No. 4 (1961), 39-45.
4. Robert A. Feldmesser, *The Option: An Analysis of an Educational Innovation*. Dartmouth College Monograph, 1969.

¹²The pass/fail option may increase the attainable set if the field of choice is defined in terms of grades.

¹³Other studies on pass/fail reveal similar results. However, Feldmesser's methodology appears to be more satisfactory since he concealed the intent of the study and considered a student's entire bundle of courses as well as attaining data on student time expenditures in various courses.

¹⁴Advocates of pass/fail implicitly argue that the option course is treated like an inferior good because the option is restricted and therefore changes relative prices. The inference is that a universal system would have income effects only.

¹⁵Note that it is conceivable that these new techniques may decrease the attainable set. In this case, a student may be forced to spend more time in the required course to maintain a passing grade at the expense of sacrificed achievement in other courses (e.g., major).

Research in Economic Education: Are Our Horizons Too Narrow?

Judith Yates

1. Introduction

There has been an increased awareness in recent research literature that the earlier experiments in economic education have been limited in both scope and experimental design. Two directions for further research have been suggested. The first suggestion has been to improve the technical aspects of evaluation. An example of this is Soper [15, p. 40], who claims that "research in economic education at the collegiate level has now progressed to the point where professionally acceptable results require the use of normed, reliable and validated instruments (such as the TUCE) and O.L.S. regression models" and that "this field of research has now reached a level where increasing sophistication in both theoretical specification of models and empirical techniques is necessary if further advances are to be made."¹ The second suggestion has been to extend the scope of this evaluation. An example of this is van Metre [16, p. 101], who argues that the first approach is too limited and that "economic educators should stop looking for the best economics teaching method and, instead, look for flexible learning systems which can embody appropriate teaching methods as indicated by the type of learning outcomes desired and the learner characteristics of the particular students involved."²

The aim of this paper is to point out that if the only direction taken by Soper's so-called second generation research is the use of second generation econometrics, then the problem of "poorly drawn conclusions and inappropriate experimental design" [16, p. 95] may be exacerbated. The desire for "professionally acceptable results" has already led to a concentration on measures of output for which "normed, reliable and validated instruments" are available. As a result of this, conclusions regarding the efficacy and desirability of the teaching methods in question have been made on the basis of these readily available and easily measured outputs. Less easily measured outputs have been ignored. This has occurred even though one method of teaching can only be described as better than another if it is possible to assess all the objectives of a course unambiguously and to decide what the weight of each objective should be. Implicit in the conclusions of much of our research data is that the major criterion or objective of economic education is a very narrowly defined concept of learning, related in some way to student performance as measured by achievement test or course grade.

In the following sections it is argued that student performance is not the only criterion which could be used in evaluation and that once the possibility of multiple criteria is considered, there will be difficulties in assigning them weights. The conclusion is that there is a danger of our research horizons becoming too narrow because of these inherent difficulties, and it is suggested that this possibility can be avoided by a conscious attempt to encourage research in both and not just one of the two directions suggested.

Dr. Judy Yates is a lecturer in economics at the University of Sydney.

2. A Single Criterion for Evaluation?

Before considering the question of what the major criteria should be for evaluating a particular teaching method used in economic education, it is important to consider the much broader question of the objectives of education in general. Johnson [9] has argued that the objectives of higher education can be to produce human capital, to serve as a source of knowledge, or to act as a filter for the labor market. Although an individual course of itself cannot bear the burden of all of the objectives of the whole education system, it must surely be recognized that the course is just a part of this whole system and cannot, therefore, be treated in isolation. The contribution that a course makes toward the objectives of the education system must always be borne in mind. These objectives are often ill-defined and may vary significantly among institutions (for example, between universities and colleges). At the risk of overlooking important ones, a number of objectives can be listed:³ students' growth and development with regard to knowledge and skills (which can include application, critical thinking, creativity, motor skills, etc., as well as comprehension and understanding); their social development (including leadership, communication, interpersonal relations, etc.); their acquisition of vocational skills, and so on. Many of these skills will not be measured in any way by the use of a TUCE (*Test of Understanding in College Economics*) type of test; nor will they be developed by many of our approaches to education. Concentration on formal learning and the related skills of analysis quite possibly arises from a questionable belief that those qualities can be objectively assessed.⁴ However, such concentration may also result from treating a specific course in isolation rather than as a part of the whole education process.

Objectives such as an increase in knowledge and/or skills at varying degrees of sophistication which are evaluated by achievement or course grade are obviously part of what is being sought. However, the fact that they continue as major objectives despite the evidence of lack of anything like perfect retention one or two years later⁵ could be taken to indicate that this specific knowledge is, of itself, not the sole objective of the course in question but that it is possibly being used as a proxy for some other objective. It may be, for example, that what needs to be considered and what is implicit in our choice of output measure is the *process* by which this output is achieved rather than the output itself.⁶ The contribution of a particular technique or teaching method to the process of learning how to learn rather than to the specific output of precisely what was learned may be its most important attribute. Allison [2] has suggested that learning how to learn does take place from one semester to the next and that the extent of this process might vary according to the teaching method employed. If the process is regarded as being as important as the output, there may be a need for more subtle evaluation measures which can distinguish between the contribution of a particular method to the process of learning and its contribution to the outcome of learning. In some teaching methods students are required to organize the material for themselves; in others it is done for them. In some teaching methods students have the discipline of steady progress through the subject imposed upon them; in others, they are left to their own resources. These teaching methods may reveal no apparent differences in their effect on achievement, but if either the ability to organize one's own material or the self-discipline required to organize one's self is desired as an objective in its own right then the effect that the choice of teaching method has on the possibility of attaining these must also be taken into account.

To illustrate the potential weakness of basing conclusions on consideration of just one single objective, Soper's own work can be cited.⁷ His and Thornton's fairly thorough study of a Keller-plan approach to economic instruction [14, p. 89] claims to have shown "that a completely self-paced teaching format for macroeconomics is inferior to a well-directed concept-oriented, graduate-student instructed, lecture-discussion taught course." This conclusion is based on no discernible improvement in grades and is made despite a long list of potential criteria which were not considered but by which such an approach could have been

assessed. As an indication of some of these criteria, in addition to the ones already raised, those designated by Allison [1] could be mentioned. She lists as benefits the fact that a self-paced system has a nonthreatening system of assessment and hence a resultant reduction in student anxiety; that it results in increased student control over grades; that there is a strong preference for this method over the conventional alternative; that it has a virtual guarantee that *all* students will achieve mastery of the subject; and, finally, that it is exciting for the instructor to introduce and conduct a new teaching method which is so favorably received by students. In addition, in a later paper [2] she also suggests that the organization of material which occurs in a self-paced approach may help students in learning the how-to-learn process discussed earlier, particularly in the early stages of higher education. To offset these potential benefits, such a system might have an adverse effect on organizational ability and on self-discipline. All these potential benefits and costs are possible objectives of a course and even if they are not necessarily primary objectives, consideration of them may well weaken the unambiguous conclusion reached by not taking them into account.

3. Which Criteria?

It follows from section 2 that the problem of choosing between two alternative approaches to teaching should not be the simple procedure implied by much of the current research. The problems of evaluation of a particular course are magnified once this course is seen as a part of the whole education system, since its evaluation cannot really be independent of an evaluation of that whole system. There will always be the difficulty of deciding on exactly how much consideration should be given to those objectives that are system-oriented compared with those that are subject-oriented. The difficulty in assigning these weights results in part from the ill-defined nature of many of the objectives, in part from a lack of consensus about them,⁸ and in part from the conflict between students and staff in their attitudes toward those objectives.

It is not improbable that students' attitudes are such that they develop the skills of surviving the system with a minimum of effort at the expense of those skills more related to a disciplined search for knowledge.⁹ Kelly [10] has used this possibility of conflict between multiple objectives to explain his seemingly contradictory findings that, although the introduction of a specific technology (namely, the distribution of printed lecture notes) had a negative impact on achievement, students were prepared to pay a positive price for this technology. He argued that this finding can be explained by recognition of the impact of the particular technology on the allocation of the students' time and by recognition of the distinction and trade-off between "educational outputs" and "leisure outputs." He concluded that "policy implications relating techniques to increased educational efficiency are clearly sensitive to a specification of *whose* values are used to identify and price the outputs associated with the instructional process."

Dual or multiple objectives at the system level, such as the teaching and research roles of a higher education institution, also have implications for the evaluation of experiments in economic education. For example, the questions of whether research or teaching is more desirable and whether there is a trade-off between them or whether they are complementary¹⁰ are of fundamental importance. They must be answered before choices can be made between teaching techniques that have different implications for staff involvement regardless of whether the techniques have different effects on the other objectives being considered.

A final and important point concerning the weighting of objectives can be made by reference to the distribution of outcomes. Many studies have shown that sex, age, ability, effort, etc., are all determinants of outcome (again, usually achievement, but occasionally attitudes or enjoyment) and fewer, but still a significant number, have actually considered the differential impact of various techniques according to these variables. Those studies which have considered distributional impact have shown that the same factors that determine outcome

are also the major factors that need to be considered when looking at the distributional effect.¹¹ Given that the outcomes of a particular approach can vary according to sex, or ability, or whatever, the decision to choose one method over another becomes an ideological matter. Unless explicit value judgments are made as to what the desirable distribution should be, it becomes impossible to decide which of two technologies that may have identical average outcomes but different distributional effects is "better." Conversely, of course, any such choice does reflect implied judgments, whether intended or not.

This distributional aspect is also relevant in considering the impact of various technologies on a student's rate and method of learning. The point will not be developed here because the fact that students do have different rates of learning has been well documented and taken into account explicitly by the various attempts at introducing self-paced courses. The second aspect, that students have varying methods of learning and that this may be an explanation of the distributional effects, has been argued more fully by the author [17] and by van Metre [16], but has possibly not received as much attention in economic literature as it ought.¹²

4. Conclusion

In conclusion, it is claimed that economic education will almost certainly have many objectives, some of which may be conflicting and some complementary. Different approaches to education will have different effects on each of the objectives both in terms of outcomes and in the distribution of those outcomes. Continued research in the direction suggested by Soper will lead to increased information about the relative efficiency and distributional impact of various approaches with respect to just one of the many possible objectives. It will provide valuable information, but it will not, and should not, provide an adequate base on which to make "best" choices. Narrow horizons, chosen simply because they lead to "professionally acceptable results," must be avoided. Information is needed about the relative effectiveness of each of the various approaches on *all* of the possible objectives, even if readily acceptable instruments are not available. This information and a knowledge of the likely distributional effects of the available alternatives will enable educators to make their own considered value judgments as to what they think will be the most effective means of achieving stated or even unstated objectives.

In the absence of a consensus concerning the desirability and importance of certain objectives, the suggestion that we should be looking toward the adoption of more flexible learning systems (including assessment systems¹³) appears to be soundly based, since if this is done there is less danger of too much weight being placed on a single outcome or a single distribution of outcomes.

FOOTNOTES

¹He argues that the use of nonrigorous theoretical and empirical procedures may have resulted in incorrect evaluations of particular educational approaches.

²His argument is essentially based on the claim that any particular course will have multiple objectives and that these are best achieved by different techniques for different students because different students have different methods of learning. As was pointed out by the author in an earlier paper [17], these same conclusions, based on very similar reasoning, were suggested by Bligh [3] and were supported by reference to a very extensive range of education research literature covering almost every field of study.

³These have been taken from [13].

⁴Increasing doubts have arisen over the objectivity of the measures used. Kipps et al. [10], for example, concluded that the specification of the dependent variable (that is, whether as absolute achievement, absolute improvement, percentage improvement, or gap closing score) could bias the results of any regression equation and so warned that "any recommendation regarding teaching methods based on such statistical findings, therefore, might be unwarranted." Unless the instructor is able to unequivocally claim that he is aiming to improve the gap-closing score, or the absolute improvement, or whatever over the whole class, then Kipp's results make interpretation of existing findings less objective than originally

thought. Likewise, Allison [2] points out that single-equation models can only provide "information on achievement, and not on other outputs such as enjoyment or concentration" and concludes that it is quite possible to introduce bias into the coefficient estimates by ignoring any simultaneity of these factors, hence making them an unreliable base for conclusions or recommendations. On similar grounds, which appeared in the Fall 1976 issue of this journal, the Soper-Becker-Highsmith exchange on research methodology by raising problems of multicollinearity, stability of coefficients between groups, measurement errors, and errors in specification, clearly points out the difficulties inherent in the search for an objective measure.

⁸Craig and O'Neill [5] suggest that the decay factor is approximately 25 percent for economics over this period, a higher figure than that suggested in other studies in economics they mention, but considerably better than the 94 percent loss they report for science education.

⁹The possibility that the process may be important is raised in [6].

¹⁰Soper's study is, of course, not the only example available. A multitude of studies have reached essentially the same conclusions on the basis of essentially the same types of results.

¹¹Horton and Weidenaar [8] conducted a survey among some two hundred economists in attempt to isolate a single goal for economic education. However, despite their use of the iterative Delphi procedure they were unable to determine agreement on a single goal. They argued that this was necessary because "without such a single goal we disperse our efforts, confuse them, and develop no solid base for the specification of instructional objectives and no systematic one for the evaluation of our efforts." Although I recognize the value of their attempts, I do not agree with their rationale.

¹²This may be a result of a conflict between the objectives of higher education of serving both as a filter for the labor market and as a source of knowledge.

¹³See, for example, McDonough and Kannenberg [12] for one particular aspect of this discussion.

¹⁴Hanson, Kelley, and Weisbrod [7] illustrate the point being made here with the aid of a simple, borrowed diagram based on only one characteristic (ability) and one output criterion (course grade). They show that it is possible for two methods to be equally effective on average but for one to favor high achievers and the other low achievers.

¹⁵It does receive considerably more attention in education literature.

¹⁶There has been wide discussion concerning the importance of the right choice of assessment method because the choice made affects the perceived achievement of the stated (and even unstated) aims of the course and determines whether the intended aims actually will be the ones assessed. See, for example, [4] for a survey of this discussion. Van Metre [16] adds several examples relevant specifically to economic education. This same point (that is, a plea for flexibility) must surely apply to the evaluation instruments being used as well; tests such as those mentioned by Soper have been shown to test knowledge and understanding and simple and complex analytical skills but whether such tests, nationally validated or not, can be used to test creativity, maturation and so on is a very debatable question.

REFERENCES

1. Allison, E. "Self-Paced Instruction: A Review." *JEE* 7 (Fall 1975).
2. ———. "Educational Production Function for an Introductory Economics Course." Discussion Paper No. 545, Harvard Institute for Economic Research, April 1977.
3. Bligh, D. "What's the Use of Lecturers?" Harmondsworth, England: Penguin, 1972.
4. Cox, R. "Examination and Higher Education: A Survey of the Literature." *Universities Quarterly*, 1967.
5. Craig, E. D., and J. B. O'Neill. "The Predictability of Retention in Economics." *JEE* 5 (Spring 1974).
6. Draayer, G. F. "Toward Program Accountability: A Model Study." *JEE* 5 (Spring 1974).
7. Hansen, W. L., and A. C. Weisbrod. "Economic Efficiency and the Distribution of Benefits from College Instruction" *Papers and Proceedings of the American Economic Review* 60 (1970).
8. Horton, R. V., and D. J. Weidenaar. "Wherefore Economics Education?" *JEE* 7 (Fall 1975).
9. Johnson, H. G. "The University and the Social Welfare: A Taxonomic Exercise." In K. G. Lumsden (ed.), *Efficiency in Universities*. New York: Elsevier North-Holland, 1974.
10. Kelley, A. C. "The Student as a Utility Maximizer." *JEE* 6 (Spring 1975).

11. Kipps, P. H., H. M. Wilhelm; and D. R. Hall. "A Note on the Use of Multiple Regression Analysis in Studies of Achievement in Economics." *JEE* 7 (Spring 1976).
12. McDonough, C. G., and L. C. Kannenberg. "The Microeconomic Impact of a Reduction in Faculty Teaching Loads." *Australian Economic Papers* 16 (1977).
13. Micek, S. S., A. L. Service; and Y. S. Lee. "Outcome Measures and Procedures Manual." Technical Report No. 70. National Center for Higher Education Management Systems, Western Interstate Commission for Higher Education, Boulder, Colorado, May 1975.
14. Soper, J. C., and R. M. Thornton. "Self-Paced Economics Instruction: A Large-Scale Disaggregated Evaluation." *JEE* 7 (Spring 1976).
15. Soper, J. C. "Second Generation Research in Economic Education: Problems of Specification and Interdependence." *JEE* 8 (Fall 1976).
16. Van Metre, D. "A Learning Theory of Economics Instructional Development." *JEE* 7 (Spring 1976).
17. Yates, J. "An Approach to Economics Education." In J. S. Cramer et al. (ed.) *Relevance and Precision*. Samsom/North-Holland, 1976.

ECONOMIC EFFICIENCY AND THE DISTRIBUTION OF BENEFITS FROM COLLEGE INSTRUCTION*

By W. LEE HANSEN, ALLEN C. KELLEY, and BURTON A. WEISBROD

University of Wisconsin

Economic efficiency implies an equating, at the margin, of benefits and costs. In this paper we explore a concept of "efficiency" which is broader than the usual framework and which applies to commodities and services produced and distributed largely outside the private, profit maximizing sector. An assessment of the economic efficiency of producing such a commodity requires the determination of its outputs and the valuation or weighting of these outputs.¹ Our principal point is that these weights, in turn, depend on who receives the outputs; thus, distributional issues are at the heart of economic efficiency studies involving a wide range of activities undertaken in the governmental and private, nonprofit sectors.² One of these activities—the production and distribution of college instruction in economics—illustrates well the significance of this particular approach to the analysis of economic efficiency.

We argue that in analyzing the economic efficiency of instruction, distributional issues—that is, who receives the benefits—should be considered explicitly; if not, they will necessarily be considered implicitly. The pervasive failure to include distributional issues in efficiency studies suggests an excessively narrow concept of efficiency. This is particularly inappropriate in evaluating instruction, since in education, as in most services, decisions regarding what to teach and how to teach have a strong influence on who receives the benefits.³

I

Total benefits from instruction are a function of the amounts gained by each student and of the values of each amount; these values vary among different students or types of students. Therefore,

* The authors are grateful to their colleagues, R. Andreano, F. Golladay, R. Lampman, and E. Smolensky, for comments on an earlier draft.

¹ We assume throughout that output is measured in incremental, units-added terms.

² This general point has been developed in detail by B. A. Weisbrod [17]. In his study the concept of efficiency which encompasses the traditional view of efficiency as well as the distribution of output is termed "grand efficiency."

³ This is true of most services because of the requirement that the consumer be present at the time and place of production (consumption) of the service.

aggregate benefits are a function of how the outputs are distributed among students. Symbolically, the marginal benefits, B_k , from resources employed in an instructional approach (technique and/or course content) k is:

$$(1) \quad B_k = \sum_{j=1}^n \frac{\partial q_j}{\partial d_k} \frac{\partial b_j}{\partial q_j},$$

where

q_j = the quantity of output produced by the input mix k and received by student j ; and

b_j = the value of benefits (output) accruing to student j .

Of the two partial derivatives the first is the marginal physical product of input (or input mix) k for student j , and the second indicates the valuation of the marginal product.

As expression (1) reveals, the importance (both in quantity and in value) of any particular form of output may vary with the type of student recipient. For a student planning graduate work certain course outputs may have great value, whereas these same outputs may be of slight value to the student who plans to continue no further in the field. Similarly, instruction about behavior of the stock market may contribute greatly to the knowledge of persons from disadvantaged backgrounds while adding little to the knowledge of other students.

Most studies of instructional efficiency or studies which appraise the merits of particular teaching techniques have not estimated—nor have they even considered—the impact of alternative teaching approaches (input mixes) on the distribution of outputs among students. Neither have they examined the possibility that the value of outputs varies according to the distribution of the output. In terms of the model, the subscript j has been entirely disregarded. The assumption implicit in such a simplification is either that students are a homogeneous group—each student receives the same amounts of outputs from a given course and the outputs have the same value

⁴ In practice, the summation will be over groups of students who possess roughly similar and "relevant" attributes. Parenthetically, it might be noted that if $j=1$, expression (1) represents the value of an input used to produce a private good.

THE TEACHING OF ECONOMICS

for each student—or that students should be treated as if they were a homogeneous group. Neither assumption seems reasonable, for rather obvious reasons.

This discussion has assumed implicitly that it is students' valuations that should count. There are many who feel, however, that valuations should be made by other groups, such as the faculty, the college administration, the taxpayers, or particular groups of students. Our general point—that the magnitude and distribution of benefits depends on the values assigned—is general and is invariant to whose values are applied.

In the following pages we discuss how instructional techniques and approaches influence the distribution of total output and the value of that output and how the distribution of output relates to a broad concept of economic efficiency.⁶

II

Consensus on course content or instructional technique has not been reached in economics or in general. The literature abounds with arguments for or against the "citizenship" focus [4] [5] [16] or the preparation of potential graduate students [13]. At another level we find some economists calling for a heavy mathematical orientation [2] (though whether as a "means" or as an "end" is not always clear); others advocate a decision-making framework [3]; and still others argue the merits of shifting emphasis from macro- to micro-economics [15]. And recently there has been much discussion of the impact of new instructional

⁶ Although much of college instruction is outside the public sector, our framework is relevant to the theory of the public sector which permits analytically separable allocation and distribution branches. If the allocation branch decisions affect the distribution of welfare, then it is not possible to determine what and how much of some commodity to produce unless costless lump-sum redistributions are possible, or unless the welfare function—involving the value-weights referred to in expression (1)—is known [10] [17]. The formulation in expression (1) permits the quantity of output to vary among consumers. This is by contrast with some of the literature on the pure theory of public goods, in which it is assumed that all consumers benefit equally in the sense of receiving equal quantities of output of the public good [14]. It is clear, of course, that even if all did receive equal quantities of output, the individual's valuations could differ greatly. Instruction does have a considerable public-good element; thus we might well have couched our argument simply in terms of differences in consumers' (students') valuations rather than in terms of differences both in their valuations and in the quantities of outputs. The difference in approach, however, is not substantive. Decisions will depend on the products of quantities and values—as expression (1) indicates—and our point is that these products are likely to vary significantly with the choice of course approach.

techniques, among them programmed instruction [1] [8], television [9] [12], and TIPS (Teaching Information Processing System) [6].

In all of these illustrations, either course content or instructional technique is being considered explicitly. But the distribution of benefits among types of student clientele is also very much at stake, and differences in judgments as to whom it is most "important" to teach may be at the root of much of the controversy. The issue of importance or values is reflected in the right-hand derivative ($\partial b_j / \partial q_j$) in expression (1), above.

Even if the importance of benefiting all types of students is equal or is assigned to be equal, the actual benefits (the left-hand derivative, $(\partial q_j / \partial k_i)$ in expression (1)) are not likely to be equal. It seems intuitively clear that different types of students—as defined by previous academic performance, desire for theoretical rigor, degree of social concern, family background, and the like—will benefit differentially, depending upon course content and instructional technique. A highly theoretical and mathematical formulation in the basic economics course may provide the largest benefits for students already thinking seriously about graduate work in economics, whereas a course focusing on a less formal treatment of contemporary economic problems (and employing a similar instructional technique)⁶ may benefit most those students seeking a "general" education. Still another choice of course content is likely to be most beneficial to pre-law students. Given the heterogeneity of the enrolled students and the difficulty of offering simultaneously a variety of course contents, some students are certain to receive larger outputs⁷ than will others. Yet little or no evidence exists on the strength of the linkage between the distribution of outputs and the choice of course content.⁸

What has just been said about course content also applies to instructional technique. The traditional theory of production assumes that the

⁶ This qualification is intended to hold instructional technique constant in the present discussion so as to concentrate attention on course content.

⁷ Important problems exist with respect to the definition and measurement of "outputs"—they may well be multidimensional—but these are outside the scope of this paper.

⁸ In practice it may be difficult to separate the quantity of output from its value, and, indeed, the two may be interrelated. For example, if the consumer places a high value on increments of a particular type of course content, then his attitude toward learning may be "better," with the result that he will receive a larger quantity of output. This would make expression (1) more complex, for $q_j/k = f(b_j/q_j)$. An exploration of this complexity is beyond the scope of our present efforts.

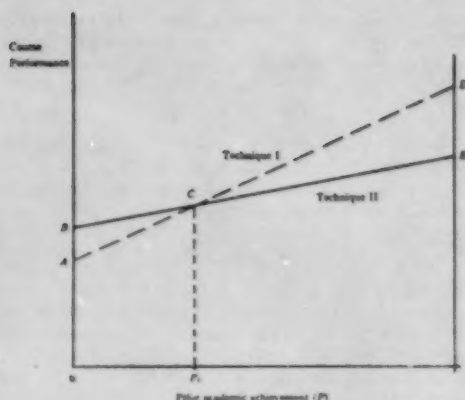


FIGURE 1

choice of production technique is a decision separable from the decision regarding the distribution of output. This assumption seems to dominate much of the evaluative research in economics education, for little effort has been devoted to finding out which kinds of students benefit by how much from the use of different instructional techniques; e.g., television and programmed instruction.⁹ The available evidence often shows whether the new technique is an improvement in the sense that students attain higher average scores than they do in courses using conventional methods. But the degree to which the new technique alters the distribution of performance is a subject scarcely ever raised.

The differential effectiveness of a given instructional technique for different groups of students and the differential value of a unit of "effectiveness" for different groups of students are illustrated in Figure 1. In this example, students are arrayed according to prior academic achievement, as measured on the X axis by a test given prior to beginning a particular course of study. The Y axis measures performance on an appropriate test of accomplishment following completion of the particular course. The curves labeled I and II show course performance—using alternative instructional techniques on two comparable groups

⁹ This is true for all subject-matter areas, not only economics, and at all levels of education, not merely collegiate. This gap in educational research has been pointed out by others. For example, Robert Locke, an executive of the McGraw-Hill Book Company, has forcefully argued, "... the ideal program should not only use the most appropriate media for each task or objective, but it should also offer alternative media to accommodate differences among learners in ability, experience, motivation, style, and rate" [7].

of students—as a function of the type of student within each group. "Course performance" reflects use of a particular test instrument for measuring "outputs" of the course.

The crossing of the two curves indicates that Technique I is more successful for the "stronger" students—those with prior academic achievement above A —whereas Technique II is more successful for the weaker students. Assuming an equal number of students at each level of prior achievement, P , it is clear that the average student gained more from the use of Technique I (that is, area CDE exceeds area ABC). Thus, if one and only one of the two techniques is to be used, the preferred approach is certainly number I. Or is it?

True, Technique I gives a larger mean and, hence, aggregate level of course performance. But what of the value of that performance? Despite the apparent superiority of Technique I, it would actually be inferior to II if the function for valuing benefits attached sufficiently greater weight to a unit benefit when realized by students with poorer prior achievement (e.g., the "disadvantaged"?). How to establish the values assigned is not an easy matter. But it is at this stage, involving valuation, that normative judgments are blended with positive findings as to the effectiveness of alternative instructional techniques or course contents.

The efficiency of Techniques I and II depends, ultimately, on costs as well as benefits. The technique producing the largest value of gross benefits (output) is not the most efficient choice if its costs are enough greater so that the value of benefits, net of costs, is smaller than for some other instructional approach. Moreover, it may be the case that none of the alternatives is efficient; that is, perhaps none produces net benefits that are positive.

Two important implications of the framework captured in Figure 1 might be noted at this point. First, the possibility of intersecting lines has profound implications regarding an analysis of the literature on the appraisal of teaching approaches. In general, although many studies in the education and economics literature show no significant impact of an experimental teaching approach, the studies are incomplete. They do not distinguish between the case of zero impact for all students and the case of positive impacts for some students and roughly equal negative impacts for others. Indeed, as McKeachie emphasizes, in his review of the voluminous literature on college teaching: "One reason for the host of experimental comparisons resulting in non-significant differences may be simply that methods optimal for some

THE TEACHING OF ECONOMICS

students are detrimental to the achievement of others" [11, p. 1157]. However, neither we nor McKeachie knows which of the two cases above is closer to the truth. The implication is that much of the prior research needs to be reworked.

Second, if there is a distributional impact of the teaching approach (i.e., if the benefit curves of the experimental and control approaches are non-parallel, whether or not they intersect), then empirical tests which omit these distributional effects will produce statistically biased results.

III

One of us has experimented with a new instructional technique in the principles course in economics at the University of Wisconsin, and some of the preliminary findings are relevant here. The details of this experiment are not important to the present discussion. What is interesting is that this experimental approach, while having a beneficial impact on all students, provided larger quantities of output (in terms of the left-hand derivative in expression (1) above) to some groups of students than to others.

Figure 2 shows that the experimental technique, by comparison with a standard lecture technique (used with an essentially randomized control group of students), produced positive amounts of output for students at every level of prior achievement.¹⁰ But it also shows that the largest outputs went to students with the lowest ACT scores.

The fact that the experimental technique dominates the standard technique may appear to suggest that our findings are uninteresting in the context of concern about economic efficiency. Since the experimental technique involves added cost, however, this is not so. Even with the observed dominance, two issues remain unresolved: (1) Are the benefits disclosed in Figure 2—the area $ABCD$ —large enough to warrant the costs? (2) If total cost is a constraint, so that the experimental technique can be provided to some students but not to all of them, to whom should it be provided?

The answer to the first question depends critically on how the outputs (benefits) are valued; i.e., on the welfare function. There is, in general, no unique value for a unit of benefit (a point of added test score), for the value is a function of who the beneficiary is.¹¹

¹⁰ In this illustration students are classified by prior level of academic achievement; clearly, however, other classifications may be appropriate; e.g., in terms of family background.

¹¹ It may also be a function of the amount of benefit

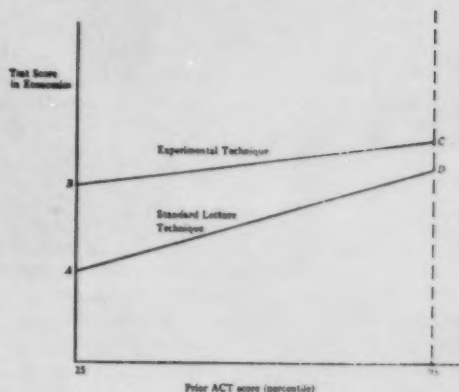


FIGURE 2

With regard to the second question, if the experimental technique is to be provided to, say, one-half of the students, which group should receive it? In other words, if total resources available are fixed, for which students will the value of output be a maximum?

The first thing to note is that Figure 2 cannot provide the answer. What it does show is that the program should be provided to the students with the lowest ACT scores if the objective is to maximize the measured quantity of output. But whether such an allocation would maximize the value of output is another issue. It will if the value of a unit increase in economics test score is a constant or decreasing function of prior ACT score, *ceteris paribus*.

If, however, there were a preference for helping "strong" students—a preference that may be reflected in the offering of "honors" courses—then it could be efficient to devote the added resources to students with high ACT scores.¹² All that is formally required to produce this result is that the value of output CD (Figure 2) exceed the value of AB , and that there be no sizable interaction effects among students.¹³ Since $AB = SCD$

realized by any given beneficiary; that is, there may be decreasing, or perhaps increasing, marginal value with respect to added units of benefit to a particular student or group of students.

¹² It would be instructive to consider the forms of weighting functions that would justify providing large amounts of resources to particular subsets of students, as is done in honors courses and in special programs for the disadvantaged.

¹³ The latter assumption is necessary to insure that the separation of students into two or more homogeneous groups would not alter their respective class performance.

in the figure, an allocation of the experimental technique toward strong students is efficient if the value of additional economics output (as defined by the test used) be at least five times as great for a student at the 95th ACT percentile as it is for a student at the 25th percentile.¹⁴

IV

The arguments in this paper have a number of research implications. Granted that an efficient allocation of instructional resources requires comparisons of benefits with costs of alternative programs, more information is needed as to the magnitudes of the benefits. These, in turn, depend on the distribution of benefits, since the value of a given absolute increment in output of instruction (i.e., achievement) is specific to the recipient. Therefore, it is clear that we need to learn more about who benefits; that is, what types of students benefit and how much each benefits when various combinations of course content and instructional technique are used. With this information in hand, normative judgments can then be applied—as expression (1) above indicates—in order to estimate the value of benefits from any particular course content or instructional technique.

The following are some specific research suggestions:

1. Batteries of pre- and post-course evaluations should be developed which reflect well-articulated output goals. While this paper has not concentrated on a precise specification of outputs, the importance of such a specification is clear. The distribution of benefits (outputs) cannot be ascertained without a prior decision—either explicit or implicit—as to the definition of output. More tests measuring different types of outputs are needed.¹⁵

2. When regression techniques or other statistical procedures are used in evaluating teaching alternatives, interaction effects should be included so as to permit estimation of relationships between the instructional approach and a variety of student attributes (e.g., class, major, mathematics knowledge, family background). Parameters of the interaction effects—such terms as the application of television or programmed learning to students with and without calculus—will provide

information about the distributional impact of the teaching approach that is being studied.¹⁶ If students benefit differentially from alternative teaching approaches, then failure to account for interaction effects between the teaching approach and student attributes will result in a mis-specified statistical model, biased parameter estimates, and often uninterpretable results.

3. Finally, it is important to face up to the issue of delineating explicitly our normative criteria on what “should” be the distribution of outputs. We must place “values” on the outputs. Until these value weights are made explicit—and the task is not simple—they will continue to be implicit and, hence, not open to critical examination.

V. Summary

The central thesis of this paper is that production decisions in education on what and how to teach have distributional effects; as a result, distributional considerations should enter directly when making teaching decisions and evaluating these decisions.

The importance of this proposition and thus the validity of the inferences derived from it involve both factual and normative matters. If alternative teaching techniques do have differential impacts by type of student (the factual issue) and/or if social objectives dictate that it is more important to benefit some types of students than others (the normative issue), then the distributional consequences of selecting a teaching technique or course approach should receive explicit attention in benefit-cost analyses of production choices.

¹⁴ With respect to the pathbreaking research of Attiyeh, Bach, and Lumsden on the impact of programmed instruction in economics [1] [8] we are encouraged to learn that analysis is currently under way to investigate some of these types of interactions, and thus the distributional effects of programmed learning.

REFERENCES

1. R. E. ATTIEYEH, G. L. BACH, AND K. G. LUMSDEN, “The Efficiency of Programmed Learning in Teaching Economics: The Results of a Nationwide Experiment,” *A.E.R.*, May, 1969, pp. 217–23.
2. D. V. T. BEAR, “The Use of Mathematics in the Teaching of the Principles Course,” in Keith G. Lumsden, ed., *New Developments in the Teaching of Economics* (Englewood Cliffs, 1967).
3. J. G. C. BOOR, “Management Games,” in Keith G. Lumsden, ed., *New Developments in the Teaching of Economics*, *op. cit.*
4. H. S. ELLIS, “This is Economics,” *A.E.R.*, May, 1961, pp. 571–78.
5. R. FELS, “A New Test of Understanding in College Economics,” *A.E.R.*, May, 1967, pp. 660–66.

¹⁴ For this problem—with the cost level constant—only an index of relative values is needed, not absolute monetary values.

¹⁵ TUCE (Test of Understanding in College Economics) and TEU (Tests of Economic Understanding) represent an excellent beginning on this long-term project [5] [18].

THE TEACHING OF ECONOMICS

6. A. C. KELLEY, "An Experiment with TIPS: A Computer-Aided Instructional System for Undergraduate Education," *A.E.R.*, May, 1968, pp. 446-57.
7. ROBERT W. LOCKE, "Educational Technology and the Educational Publisher," *Educational Technology*, Jan., 1968, pp. 14-16.
8. K. G. LUMSDEN, "The Effectiveness of Programmed Learning in Elementary Economics," *A.E.R.*, May, 1967, pp. 652-59.
9. CAMPBELL R. MCCONNELL, "An Experiment with Television in the Elementary Course," *A.E.R.*, May, 1968, pp. 469-82.
10. MARTIN C. MCGUIRE AND HENRY AARON, "Efficiency and Equity in the Optimal Supply of a Public Good," *Rev. of Econ. and Statis.*, Feb., 1969, pp. 31-39.
11. W. J. McKEACHIE, "Research on Teaching at the College and University Level," in N. L. Gage, ed., *Handbook of Research on Teaching* (Chicago, 1963).
12. DONALD W. PADEN, "Television as an Instrument of Instruction," in J. G. Lumsden, ed., *New Developments in the Teaching of Economics*, *op. cit.*
13. R. RUGGLES, "Relation of the Undergraduate Major to Graduate Economics," *A.E.R.*, May, 1962, pp. 483-89.
14. P. SAMUELSON, "The Pure Theory of Public Expenditure," *Rev. of Econ. and Statis.*, Nov., 1954, pp. 387-89.
15. G. J. STIGLER, "Elementary Economics Education," *A.E.R.*, May, 1963, pp. 633-59.
16. L. E. WAGNER, "Task Force to Classroom," *A.E.R.*, May, 1963, pp. 660-73.
17. B. A. WEISBROD, "Income Redistribution Effects and Benefit-Cost Analysis of Government Expenditure Programs," in Samuel B. Chase, Jr., ed., *Problems in Public Expenditure Analysis* (Brookings, 1968).
18. A. L. WELSH AND R. FELS, "Performance on the New Test of Understanding in College Economics," *A.E.R.*, May, 1969, pp. 224-38.

III
Foundations of Research

CONCEPTUAL AND EMPIRICAL ISSUES IN THE ESTIMATION OF EDUCATIONAL PRODUCTION FUNCTIONS*

ERIC A. HANUSHEK

ABSTRACT

Measuring educational performance and understanding its determinants are important for designing policies with respect to such varying issues as teacher accountability, educational finance systems, and school integration. Unfortunately, past analyses of student achievement and educational production relationships have been plagued by both a lack of conceptual clarity and a number of potentially severe analytical problems. As a result, there is considerable confusion not only about what has been learned, but also about how such studies should be conducted and what can be learned. This review considers each of these issues and also relates knowledge from these studies to research about areas other than just school operations and performance.

Despite a substantial and growing volume of research into educational production relationships and the determination of student achievement, considerable confusion remains about how such studies should be conducted, how past analyses should be interpreted, and what has been and can be learned from such studies. These studies are interesting and important from a number of perspectives. First, they exemplify the difficulty frequently encountered in the empirical application of some basic economic models, and the lessons there apply to a wider class of problems than just understanding schools. Second, the results of these studies have ramifications for a variety of analyses that do not focus on schools, such as wage determination, status achievement, the financing of schools, and the impacts of quality of education on urban location and housing choice.

The author is Professor of Economics and Political Science, and Director, Public Policy Analysis Program, University of Rochester.

* This paper benefited from helpful comments on previous versions by Charles Bidwell, Mary Curran, Henry Levin, Richard Murnane, Edward Pauly, John Quigley, James Rosenbaum, and especially Richard Nelson. The research was largely done while the author was at Yale University. [Manuscript received May 1978; accepted November 1978].

The Journal of Human Resources • XIV •

However, perhaps most important—and a distinguishing feature of this research—is that the concluding sections of these papers on “policy implications” do not have the more common hollow ring because the results of this research have frequently entered into judicial proceedings, legislative debate, and executive branch policy deliberations. Unfortunately, this research has been unusually difficult to follow, in part because its development has frequently crossed disciplinary boundaries.

I. SOME BACKGROUND

The first, and perhaps still the most influential, study is *Equality of Educational Opportunity*, or the “Coleman Report” [37]. That study, mandated by the Civil Rights Act of 1964, was startling in a number of ways. First, the survey information for over half a million students, containing data not only about students and the characteristics of their more than 3,000 schools, but also about their achievement in school, provided the most complete description of elementary and secondary schools ever produced in this country. Second, and most relevant for this discussion, it directed attention to the importance of the relationship between school inputs and student achievement.¹ Finally, it introduced into the public policy arena a bewildering array of technical and esoteric issues such as statistical significance, analysis of covariance, production efficiency, multicollinearity, residual variation, estimation bias, and simultaneous equations.

The attention paid to the input-output analysis in the Coleman Report clearly reflects the direct policy importance of the analysis. Such information is critical not only to “school management,” but also to such diverse policy issues as school integration, accountability in schools, and the finance of elementary and secondary schools. The policy relevance of input-output studies has led to both a rapid growth in number of analyses and a concerted effort to interpret the many different, and apparently contradictory, results (e.g., Hanushek and Kain [66], Bowles [21], Bowles and Levin [23, 24], Cain and Watts [28], Averch and others [11], Levin [91]).

As economists entered this area, the relationships estimated became known as “educational production functions” instead of simply input-output analyses. This was more than a semantic change, however; the term *production function* has special connotations that alter the interpretation of the results. In fact, one part of the discussion of these analyses is whether or not the models estimated are production functions in the usual technical meaning of the term.

1 This perspective on the analysis of schooling, as pointed out in the response of Coleman [38] to his critics, was perhaps the most important aspect of the “Coleman Report.”

II. TEXTBOOK ANALYSES

All sophomore economics classes develop the concept of a production function and its uses in analyzing the decisions of firms. In its usual abstract presentation, a firm's production possibilities are assumed to be governed by certain technical relationships, and the production function simply describes the maximum output feasible with different sets of inputs. The key distinction between a "production function" and any number of alternative descriptions of input and output relationships is the notion that it represents the maximum achievable output for given inputs. Firms are conceptualized as attempting to maximize profits through decisions about level of production and mix of inputs (given the production function, product demand, and input prices). With some embellishments, this represents the motivation for and theoretical backdrop to production function studies.²

The production function, along with the related theoretical apparatus of optimal firm decisions, is a powerful pedagogical tool, since it provides a basis for describing efficient production, the appropriate response of firms to changes in technology or input costs, and so forth. Further, the basic analytical constructs seem applicable to a wide variety of applications—there is *a priori* no indication that this structure applies to, say, the steel industry and not the education industry.

These theoretical concepts are, however, deceptively simple. When taken out of the classroom, they often require substantial modification. At one level, the concept of a technologically based production function may, in reality, not be particularly applicable in a variety of actual situations.³ At another level, even if appropriate, the actual production function is generally

2 This analysis can be extended to consider multiple outputs or wider ranges of inputs. Additionally, certain assumptions about the form of the production relationships can be included. For example, declining marginal products for inputs are often assumed; that is, the additional output per unit from increasing a given input might be expected to decline as larger quantities of the input are used (holding constant other inputs). One might also believe that certain inputs are complements in production; that is, the value of a given input might increase as more of another (complementary) input is used. Finally, the possible effects of different scales of operation can be incorporated.

3 The standard "textbook" view is that production functions are derived from known engineering relationships that reflect exogenously given technological processes. The firm decides upon a mix of inputs, and the best process for combining these inputs is indicated by the production function and, therefore, does not have to be explicitly considered. A recurring theme, elaborated below, is that this may not be a good characterization of many production processes and firm decisions where: (1) individuals involved in production actually have considerable discretion in choice of process; (2) the "best" process might not be generally known and uncertainty is important; or (3) dynamics are important and the production technology is changing. For a general statement of these issues, see Nelson and Winter [109, 110, 111].

not known a priori and must be estimated based upon the observed operations of firms. However, such estimation raises a series of conceptual and statistical problems.

Although some significant differences exist in the application of production functions to education and to other industries, the biggest differences probably arise from the potential uses of the analysis. Few people would expect manufacturing firms to change their behavior given estimated production functions for manufacturing industries (see, for example, Hildebrand and Liu [71] or Griliches [54]), and there is very little temptation to prescribe any public policies based upon the results. The same cannot be said for education. Congress holds hearings on the size of estimated coefficients (see, for example, [141]), commissions include results in supporting their policies (for example, [47] or [116]); and courts receive testimony about regression equations.⁴ Because the findings and interpretations of educational production functions go far beyond answering a set of esoteric questions of economists, educational production functions have been discussed more widely, and the confusion surrounding them seems somewhat greater, than production function estimation in general.

III. CONCEPTUAL PRODUCTION FUNCTIONS AND EDUCATIONAL REALITIES

Studies included under the rubric educational production functions are generally statistical analyses relating observed student outcomes to characteristics of the students, their families, and other students in the school, as well as characteristics of schools. Most frequently, student outcomes are measured by various standardized test scores, although attitudes, college continuation, and attendance patterns have also been analyzed. These studies also diverge considerably in terms of the actual measured inputs; in terms of the level of aggregation of both dependent and independent variables (e.g., individual student, school average, or district average observations); and in terms of the precise statistical methods. Not surprisingly, given such differences, the conclusions of the various studies appear to be very different—and often apparently contradictory.

This paper considers the major conceptual and empirical issues in such analyses with an emphasis upon the implicit assumptions and alternative interpretations of past models and results. While there is no attempt to review systematically individual studies, a later section summarizes the

4 For example, *Keyes v. School District No. 1*, 413 U.S. 189 (1973); *Serrano v. Priest*, 5 Cal.3d 584 (1971); *Hobson v. Hansen*, 269 F.Supp. 401 (D.D.C.1967).

major findings and ambiguities along with indicating some areas of profitable future research.⁵

A. Measurement of Output

The vast majority of production function studies measure output by standardized achievement test scores. A few have also considered other measures such as student attitudes (Levin [89f, Michelson [101], Boardman, Davis, and Sanday [18]), attendance rates (Katzman [80]), and college continuation or dropout rates (Katzman [80], Burkhead, Fox, and Holland [26]). Are any, all, or none of these sensible measures of educational output?

While standard production theory concentrates upon varying quantities of a homogeneous output, this is not easily translated into an educational equivalent. Education is a service which transforms fixed quantities of inputs (i.e., individuals) into individuals with different quality attributes. Educational studies rightfully concentrate upon "quality" differences. However, simply because individuals can be ordinally ranked in terms of cognitive test scores does not imply that such a measure is necessarily appropriate.

Perhaps the most important concern with standardized tests is the lack of external validation. These tests do discriminate among individuals; that is, they can divide the population into different groups. However, questions are generally selected by criteria internal to tests: (a) their ability to divide students (so that questions that can be answered by all or none of the relevant population aren't useful); and (b) their consistency with other questions (i.e., whether individuals getting a given question right tend to get other questions on the test right). Further, a given test should produce the same score if taken at different times by the same individual, and slightly different wordings of questions covering the same concept should yield the same results. None of these relates directly to whether or not tests cover material, knowledge, or skills valued by society.

Clearly, much of the observed interest in school system performance relates to the perceived importance of schooling to future capabilities of students—the ability of students to cope with and perform in society *after* they have left school. To be sure, there is some value to knowledge for its own sake, other things being equal, and more knowledgeable individuals

⁵ Perhaps the most thorough review of the findings *per se* is still Averch and others [11]. However, this is now somewhat out of date. A number of noteworthy studies do not appear in that volume: Central Advisory Council [32], Levy [92], Boardman, Davis, and Sanday [18], Garner [50], Perl [115], Heim and Perl [69], Summers and Wolfe [133, 134], Murnane [106], Winkler [149], Jencks and Brown [77], Henderson, Mieszkowski, and Sauvegeau [70], Armor and others [6], Ritzen and Winkler [120], Winkler [150], Maynard and Crawford [100], Link and Ratledge [96].

may be more interesting, happier, or whatever. However, if schools were perceived to perform a simple monastic role, it is inconceivable that they would receive the attention and interest that they do. Here we consider two dimensions of school effects: the effect on labor market performance and the effect on socialization—that is, political awareness, citizenship, moral values, etc.

Economists have analyzed the influence of education on earnings and labor market performance (see reviews by Mincer [103] and Rosen [123]). Sociologists have explored the effects of schooling on occupational choice, mobility, earnings, and the relationship between schooling and personal and family characteristics (Sewell and Hauser [126], Boudon [20], Jencks and others [76], Alwin [4], Duncan, Featherman, and Duncan [44], Jencks and Brown [77], Blau and Duncan [17]). These studies direct attention to the critical question of what role formal education plays in influencing later lives of citizens—a focus frequently lost in research into school operations.

However, a recurring problem with such studies is the inadequate measure of the education individuals receive. Most commonly, years of schooling is used to measure education. (This is even the case in models of “human capital production functions”; see Ben-Porath [13].) Few measures of the quality of education have been incorporated in such studies. Since the most pressing school policy questions concern how to improve the quality of education, this is a particularly unfortunate limitation.

Some attempts have been made to incorporate qualitative measures, such as information about cognitive abilities of individuals or about school expenditure levels into labor market studies.⁶ Such studies have been severely limited by data availability, the necessity to use fairly peculiar samples, and stringent assumptions about school operations. (For example, expenditure studies *assume* expenditure differences index quality differences.) Futher, the results with respect to the effects of quality differences have been quite inconclusive. Thus, while these studies offer an important perspective on how to observe educational outcomes, they do not currently provide much guidance to studies focusing on the operations of schools.⁷

6 See Welch [145, 147], Weisbrod and Karpoff [143], Ashenfelter and Mooney [10], Rogers [122], Weiss [144], Hansen, Weisbrod and Scanlon [60], Hanushek [62, 64], Johnson and Stafford [78], Morgenstern [104], Taubman and Wales [136], Solmon [129], Link and Ratledge [95], Jencks and Brown [77], Ribich and Murphy [117], Lee [87], Wachtel [142], and Akin and Garfinkel [2]. Only the Welch studies and the Jencks and Brown study attempt to consider the operations of schools.

Research on “ability” and earnings (e.g., Griliches and Mason [56] or the review in Griliches [55]) is also related if ability is considered endogenous.

7 Related research into evaluating the effects of various manpower and job-training programs on subsequent labor market performance (e.g., Cain and Hollister [27], Ashenfelter [9], Kerachsky and Mallar [81], and Kiefer [82]) offer a similar perspective.

Although the relationship of schooling and labor market performance is central to many policy questions, it is not the only area of interest. Hence, studies have also examined the role of education in increasing job satisfaction (Duncan [43], Black [16]), in maintaining personal health (Grossman [57], Manheim [97]), and in increasing the productivity of mothers engaged in household production, as well as the effects of the mother's education on the learning of young children (Hill and Stafford [72], Leibowitz [93], Lindert [94], Inman [73]). Further, political scientists have considered the effect of education on political socialization and voting behavior (Campbell and others [29], or the review by Niemi and Sobieszek [112]), and sociologists have considered the relationship between education and criminality.

While these studies have suggested some gross effects of quantity of schooling on other life outcomes, they virtually have never addressed the question here—how such outcomes vary in response to differences in school programs and operations. Analysis of nonlabor market areas—if refocused toward the performance of the educational system and qualitative differences in schooling—does have the potential for providing a more balanced perspective on educational productivity. Nevertheless, existing studies have yielded inconclusive results about the effects of even quantity of schooling, let alone the more detailed information.

A more fundamental shortcoming is the superficiality of the conceptual notions of the mechanisms by which education affects skills and later experiences. Cognitive skills, the chief measure of educational quality, may not be the only, let alone the most important, outcome of schooling in determining individuals' future success. One might think that more educated individuals can accomplish given tasks better or more swiftly, but surely this holds for only certain types of jobs. Less education may even be better in jobs requiring manual skills or jobs that are very repetitive. One rather commonly held presumption is that better educated individuals are able to perform more complicated tasks or are able to adapt to changing conditions and tasks (see Welch [146] and Nelson and Phelps [108]). This hypothesis has important implications for studying the productivity and outputs of schools through understanding of the mechanisms by which school interacts with the work place. Such understanding could provide considerable insight into how to measure the outcomes of schooling (or at least where to look) and how these outcomes might change with the character of the economy. The lack of conceptual clarity holds equally for other potential outcomes of the educational system.

While these studies have in general not considered programmatic differences in any detail, such consideration could provide additional insights into the characteristics of educational programs that are important for future success.

The uncertainty about the source of schooling-earnings relationships is also highlighted by the recent attention to "screening" aspects of schooling. Schools may produce more qualified individuals or may simply identify the more able. The latter view has been the subject of both theoretical and empirical treatment by economists and sociologists (Spence [130], Taubman and Wales [136], Berg [14], Thurow [138], Thurow and Lucas [139], Riley [119], Arrow [7], Stiglitz [131], Wolpin [152], Layard and Psacharopoulos [86]). Most of the attention paid to screening models arises from the implication that the social value of schooling may be considerably less than the private value (that is observed in earnings relationships) if schools are merely identifying the more able instead of actually changing their skills. Further, the screening model suggests both possible reinterpretation of the historical contribution of education to economic growth (see Denison [41]) and revisions of expectations about future returns to schooling. (These revisions depend upon the "quality" of the screening function as schooling distributions change and the response of firms to any such changes.) However, there are also direct implications of the screening model for the measurement of educational outcomes and the analysis of educational production relationships. In a screening model, the output of schools is information about the relative abilities of students, and this would suggest that more attention should be directed toward the distribution of observed educational outcomes (instead of simply the mean outcomes) and their relationship to the distribution of underlying abilities. Further, the interpretation of some studies, such as those of school dropout rates discussed below, might be radically altered, since schools with a higher dropout rate might actually be providing better information (higher output) than those with lower rates—an interpretation that is very different from that of the authors of these studies. Unfortunately, no persuasive test has been devised to distinguish between a screening model and the more standard "production" model.

These two views—production and screening—are also not the only models explaining subsequent performance. For example, Jencks and others [76] argue that luck and personal characteristics (that are unrelated to schooling) are the most important determinants of earnings differences. Bowles and Gintis [22] believe that earnings differences arise chiefly from the existing social structure and that schools adjust to instead of determine subsequent outcomes. While these latter two views are not completely convincing, available evidence does not conclusively differentiate among these four divergent views.⁸

8 The Jencks and others [76] conclusions rest upon the finding that regression analyses of earnings include sizable unexplained variation which they label "luck." However, since these analyses are based upon a small number of crudely measured characteristics (years of

Referring back to the original question, we find simply a large degree of uncertainty about the appropriateness of test scores as outcome measures. While the various studies of lifetime outcomes are conceptually very relevant to measuring school outputs, they have not been particularly illuminating for the study of school production functions. While it would not be particularly surprising if standardized test performance was not highly correlated with future success (since the standard test construction methodology makes little pretense of relating test performance to any external criteria), available empirical evidence is inconclusive about whether or not there is some fortuitous linkage between test scores and subsequent achievement.

Nevertheless, performance on tests is being used to evaluate educational programs, and even to allocate funds, and there are some pragmatic arguments for the use of test scores as output measures. Besides their common availability, one argument is that test scores appear to be valued in and of themselves. To a large extent, educators tend to believe that they are important, albeit incomplete, measures of education. Further, parents and decision-makers appear to value higher test scores—at least in the absence of evidence that they are unimportant. (Note the continued pressures to make scores more publicly available.)

A more persuasive argument for the use of test scores relates to continuation in schooling. Almost all studies of earnings which include both quantity of schooling and achievement differences find significant impacts of quantity that are independent of achievement differences.⁹ This implies that achievement differences do not adequately measure all skill differences. However, at the same time, test scores appear to have an increasing use in selecting individuals for further schooling. Thus, they may relate directly to

schooling, age, and perhaps measured achievement-test scores or family background), it should not be surprising that much is left to be explained. Moreover, there is little basis for labeling our ignorance (the residuals from a regression analysis) in any particular manner. Direct analysis of individual earnings over time (Hanushek and Quigley [67]) indicates that about two-thirds of the unexplained variation in earnings models represents stable, but unmeasured, individual factors.

The Bowles and Gintis conclusions [22] in part rest on similar evidence—measures of IQ or cognitive ability differences do not explain much of the variation in individual earnings. This, combined with an analysis of the historical development of U.S. schools, is used to support their "social structure" view of earnings determination. However, their analysis relates solely to the U.S. economy and the U.S. schooling system. There is no independent analysis of how differences in social structure affect earnings possibilities, even though this is presumably the relevant evidence.

- ⁹ An exception is Hansen, Weisbrod, and Scanlon [60], but also see the comments on this study by Chiswick [33] and Masters and Ribich [98]. See also Gintis [51] for a review of some of the literature on this topic up to about 1970.

relate directly to the "real" outputs through the selection mechanism.¹⁰

Finally, a few miscellaneous issues about output measurement should be added. First, if one does use test score measurements, there are a number of choices, related simply to the scaling of scores. Tests are often available in "grade level" equivalent, percentile ranking, or raw score forms, all of which provide the same ordinal ranking (except for the possibility of some compression of the rankings). Yet, for most statistical work, one wants a scale which indicates how different individuals are rather than one that simply ranks them. The choice really depends upon the relationship of these estimates of output to the subsequent outcomes, which are not known.¹¹ Second, there is some movement toward criterion-references tests—tests that relate to some set of educational goals. The crucial issue is the development of goals. The previous discussion argues for goals that relate to performance outside of schools, but it is not obvious that these goals guide much of the current development work.

B. Multiple Outputs

Most educational production function studies have analyzed a single output or, alternatively, a series of output measures without consideration of their interactions. (Exceptions include Levin [89], Michelson [101], Boardman, Davis, and Sanday [18], and Brown and Saks [25].) If indeed the educational process is best characterized as producing a set of outcomes (say, cognitive skills and political awareness) and if there are important interactions among them in production, then interpretation of commonly estimated models for a single outcome becomes complicated.

Ordinary least squares (OLS) regression analysis, which is commonly used in analyzing production functions for single outcomes, is generally inappropriate when there are multiple outcomes that are simultaneously produced. In the simplest case, assume that there are two outputs, one of

10 This argument is found in Dugan [42]. The use for predicting future school performance and for selection is also central in Wirtz and others [151].

The interpretation of tests and their use in selection may be changing, however. On the one hand, at least by newspaper accounts, there is a growing concern about the information contained in test scores. On the other hand, courts are increasingly becoming concerned with the use of tests for selection, particularly when they might have discriminatory outcomes. For example, in *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971), and a host of similar cases, a central issue is whether test performance relates to job performance.

11 This is actually just a special case of more general questions about the functional form of production functions (discussed below). Coleman and Karweit [39] argue against use of grade-level equivalents on the basis of their peculiar properties even for answering qualitative questions. Whether or not tests measure accurately the activities and learning that take place in schools has not been considered here. For the purposes here, such concern simply deserves little attention if unrelated to subsequent outcomes.

which is an "intermediate" outcome (such as attitudes toward school) that is not valued itself, and one of which is a "final" outcome (such as achievement) that is valued by decision-makers; also, assume that the underlying production relationships (the structural equations) are such that, in addition to a series of exogenous factors (such as family background and schools), attitudes affect achievement and, simultaneously, achievement affects attitudes. In such a situation, both structural equations can generally be estimated simultaneously (although not with OLS methods). Alternatively, it is possible to estimate the "reduced-form" equation for the separate outcomes (where the reduced form is the relationship between one of the outcomes and the exogenous variables and is found by substituting one structural equation into the other). The reduced-form equation, which can generally be analyzed with OLS techniques, indicates both the direct and indirect impacts (through the other outcomes) of the exogenous variables. The reduced-form estimates do not indicate the process by which an exogenous variable affects an outcome and may be misleading if one undertakes policies that change the structural relationships. However, there are no quarrels with the underlying statistical methods or the judicious interpretation of the results.¹² Many single-equation analyses can be interpreted simply as attempts to estimate reduced-form relationships.

However, alternative multiple-outcome models are more complicated. Consider again the case of two outcomes except now let both be "final" outcomes that are independently valued by decision-makers. With information about the alternative outputs, inputs, and decision-makers' valuation of outputs, the structural equations can again be estimated directly.¹³ However, with information about only one output, estimation of the reduced form might be quite misleading.¹⁴ The estimated effects of various inputs will reflect both the production technology (the effect of each input

12 Underlying this is a series of complicated statistical arguments and assumptions. For example, there are the issues of identification, the distribution of the error terms, etc. There are also specialized forms of the structural models—such as recursive models—which can be estimated by OLS. Discussion of these issues along with discussion of the desirability and methods of estimating the structural equations can be found in Hanushek and Jackson [65, Chs. 8 and 9].

Reduced-form estimation still requires some specification and measurement of the other equations in the system, which may be difficult. For example, with attitude formation, little is known about the determinants of attitudes, and clearly experiences outside schools (for which data are often lacking) are important.

13 There is, in reality, little information about decision-makers' choices among outcomes. This makes estimation of the structural equations very difficult, even when data are available about the relevant outputs and inputs.

14 In the general case, the errors in the reduced-form equation will be correlated with the exogenous factors through the decision function about different outputs.

on the single output) *and* the choice between outputs, not simply the production technology.

The empirical importance of this issue is generally unknown. It depends importantly upon the degree of "jointness" of production, the form of the production function, the variance of choices, the underlying decision rules for determining choices, and the accuracy of measuring inputs. It is possible to construct models where joint production appears extremely important (e.g., Brown and Saks [25] develop a model where both mean and variance of achievement are valued and where simple reduced-form estimates appear quite misleading). However, there is a wide variety of circumstances where such issues are inconsequential.¹⁵ Without more information about both the range of outputs (and their measures) and the potential decision rules, there is little that can be said about this problem.

Given measures of alternative outputs, it would be appealing to look at production functions for "total" output. This is essentially what is done for production function estimates in other sectors, where market prices are used to aggregate outputs. However, these prices are not available for education. Even if available, they may be inappropriate since the weights in the decision function for outputs may differ from the market prices.¹⁶

Consideration of multiple outputs does suggest that production functions estimated with test-score measures might be more appropriate in earlier grades, where the emphasis tends to be more on basic cognitive skills—reading and arithmetic—than in later grades. In other words, these outputs appear to be much more heavily weighted than others at earlier grades, and therefore the potential problems of multiple outputs are less than in later grades.¹⁷

C. Inputs to the Production Process

The usual prescription for developing the relevant set of inputs to a produc-

- 15 Take a simple example where one is concerned with alternative outputs which are independently produced; say, one is concerned with reading ability and sex education and they do not interact (i.e., each does not appear in the other structural equation). As long as we have accurate measures of all exogenous variables (discussed below), estimation of a single equation may not be affected by the decision process which weights the two outputs. Alternatively, if the relative weights placed upon the two outputs vary dramatically—say, cognitive skills are emphasized much more than other skills—the problems may be empirically insignificant.
- 16 This is not an issue in studying competitive industries where it is assumed that managers maximize profits; thus, the weights in the output decision function are simply the market prices.
- 17 Note that all production function studies have been conducted for elementary and secondary schools. In postsecondary education, few people believe that test scores adequately measure outputs.

tion process is to find an engineer who will describe the technical characteristics and specifications of the process. When considering education, the "engineers" are usually thought to be learning theorists. Nevertheless, almost all educational analyses begin with laments about how we do not have any learning theory that is suitable for guiding input-output analyses. In reality, engineers give little guidance to the development of production functions developed for any sector. Typical production function studies, say, for manufacturing industries, incorporate two or possibly three inputs (capital, labor, and possibly education level of labor), and few would argue that engineers had much to do with this specification. This set of inputs does, however, match the pedagogical models and, while there are some minor debates about such issues as the measurement of capital, generally receives limited attention.

In education, the relatively fixed input of labor and capital (i.e., one teacher per classroom with a relatively small variance in class size) implies that this simple description of inputs could explain little. Somewhat ironically, because educational studies have attempted to provide much more detail about input differences, they have been faced with much more criticism about the specification of the inputs.¹⁸

Part of this criticism is explained by the fact that input specification has not received much attention in many past analyses. There is little conceptual clarity, and the choice of inputs seems, sometimes explicitly, to be guided more by data availability rather than any notions of conceptual desirability. For example, nowhere in the Coleman Report can one find a statement of an underlying conceptual model. At times such a model seems implied, but the statistical analyses do not seem to relate to the implied model (see Hanushek and Kain [66]).

Conceptually, a model such as equation (1) seems generally acceptable:

$$(1) \quad A_{it} = f(B_i^{(t)}, P_i^{(t)}, S_i^{(t)}, I_i)$$

where, for the i th student, A_{it} = achievement at time t ; $B_i^{(t)}$ = vector of family background influences cumulative to time t ; $P_i^{(t)}$ = vector of influences of peers cumulative to time t ; $S_i^{(t)}$ = vector of school inputs cumulative to time t ; and I_i = vector of innate abilities. The development of this model and background analyses entering it are discussed elsewhere (Hanushek [61]), and this discussion will only highlight important issues.

In the abstract, it is difficult to quarrel with this specification; controversy enters only when more detail about the definition and measurement of variables and the form of the functional relationship are introduced. The first important point is that the inputs are those that are relevant to the

18 Again, part of the attention to model specification relates to the very different purposes behind educational analyses and analyses of other sectors (see above).

individual student. Additionally, the model portrays the educational production relationship as cumulative—past inputs have some lasting effect, although the value in explaining output may diminish with more distant inputs. A corollary to this last point is that, without fairly strong assumptions about the dynamics of education—that is, the time paths of adjustment to change—the data requirements are huge.

In part to circumvent some of the data requirements (and in part because of other considerations discussed below), an alternative version of this model has sometimes been analyzed. If equation (1) holds at different points in time, say, a past period t^* , we can consider the change in achievement between t and t^* as in equation (2):

$$(2) \quad A_{it} = f^*(B_i^{(t-t^*)}, P_i^{(t-t^*)}, S_i^{(t-t^*)}, I_i, A_{it^*})$$

where the inputs are measured over the period t^* to t .¹⁹ This formulation, sometimes referred to as "value added" specification, apparently lessens the data requirements, but it does so at the expense of some additional assumptions about the relationships (discussed below).

Consider now the empirical implementation of these models. Most analyses are purely cross-sectional and include only contemporaneous measures of the inputs. No studies have adequate measures of initial endowments (or "learning capacity"). Many educational inputs (e.g., family educational inputs) are not measured directly, but instead are proxied by other observable attributes (such as socioeconomic background of the family). Little attention is given to the dynamic structure, that is, how the effects of different inputs cumulate. The relevant inputs (e.g., school factors) are often measured with considerable error.

The divergence of the conceptual model and the empirical models which have been estimated means that interpretation of the empirical results often requires a series of implicit assumptions, some of which are very dubious. The remainder of this section attempts to make explicit the most important assumptions underlying the empirical analyses.

The most consistent and obvious divergence of the empirical models from the conceptual models is the lack of measurement for innate abilities. In fact, there is little clarity about what should be measured in this term (I_i).

¹⁹ This equation results from simply subtracting equation (1) for time t^* from equation (1) for t . However, instead of analyzing $A_{it} - A_{it^*}$ as the dependent variable, A_{it^*} is put on the right-hand side. There are three reasons for doing this: (1) empirically, A_{it} and A_{it^*} may well be different tests with different scaling, etc.; (2) levels of starting achievement (A_{it^*}) may influence achievement gain; and (3) correlated errors in achievement measurement may suggest such a formulation (Cronbach and Furby [40]). However, the latter argument suggests that further corrections for errors in the exogenous variables—probably based upon test-reliability measures—are also needed since such errors, even if they have zero means, will yield inconsistent estimates.

Presumably, it includes "learning capacity," but this is not well defined.

In a regression framework, the effect of omitting an important variable is bias in the estimated regression coefficients. The importance (size of bias) is related both to the strength of the variable on achievement and the correlation of the omitted variable with other included variables in the model. If innate abilities were uncorrelated with all of the included variables, the only effect would be to increase the residual variance, and there would be no bias of the other coefficients estimated. However, there is some evidence that these correlations are not zero. If I_i is related to IQ, we know that, in particular, IQ is correlated with family background (either through genetics or environment). For example, Scarr and Weinberg [125] find that 10 to 30 percent of the variation in IQs of adolescents is explained by measured family characteristics.²⁰ Further, the correlations for younger children are higher. This implies that the omission of innate abilities probably biases upwards the estimated impact of family background on achievement.²¹ At the same time, it is plausible to assume that biases in other parts of the model will be considerably less, particularly in the case of school inputs. The correlations between innate abilities and school attributes, after allowing for family background factors, is likely to be small.²² Likewise, the importance of these omitted factors is lessened if the estimated model is equation (2), since "level" effect would be included in A_{it} and only "growth" effects of innate abilities would be omitted. (See Boardman and Murnane [19] for a discussion of potential biases in alternative specifications.)

The next major category of empirical problems is the accuracy of variable measurement, a problem which occurs in several different forms. Frequently, only contemporaneous measures of the exogenous variables are

- 20 The reported numbers are R^2 s for regression equations which include family attributes (e.g., mother's and father's education, etc.) on children's IQs in biological families. The range in R^2 s basically arises from the inclusion of parents' IQs. For adoptive families, the R^2 s range between .02 and .16.
- 21 The bias is complicated in models with many exogenous variables; in that case it depends upon the sample partial correlations of the omitted variables on all of the exogenous variables. Plausible assumptions about the partial correlations, however, indicate an upward bias in family background coefficients. For details of biases, see Hanushek and Jackson [65, Ch. 4].
- 22 Innate abilities are probably positively correlated with school attributes and peers because higher SES families generally live in relatively homogeneous neighborhoods and because they also select, or demand, higher quality schools. However, the concern is not the simple correlations, but instead the correlations after controlling for family background differences; these are likely to be considerably smaller. On the other hand, the existence of ability-tracking might increase the correlations of innate abilities and school inputs if tracked students systematically receive different school inputs (see Rosenbaum [124] and Alexander and McDill [3] on tracking and school inputs).

available, implying that the cumulative variables are generally measured with considerable error. Even if the errors of measurement have a mean of zero, the coefficients will be biased; the amount of bias is roughly proportional to the variance of the measurement error relative to the variance of the true variable (see Hanushek and Jackson [65, Ch. 10]). In this case, however, the errors of measurement for background factors (and the biases in these coefficients) are probably less than for other factors, since current measures of backgrounds give a better picture of historical factors than either current measures of peers (because of migration or changing of schools) or current measures of school inputs. These measurement error problems are undoubtedly more severe in models such as equation (1) than for those in equation (2) where the relevant history is back to t^* rather than to birth.

The biases from "historical" measurement errors are probably most severe for schooling inputs and, to a lesser extent, for peer influences.²³ Common "contemporaneous" measurement errors probably also impact most severely on school inputs. Much analysis has tried to capitalize on readily available school data—data produced routinely for administrative purposes. Typically, these data provide measures of average teacher or school characteristics, but are not linked to individual students, as called for in the conceptual model. In fact, schools are often very heterogeneous institutions with considerable intraschool variance in staff and programs. This problem becomes more acute at later grade levels where average characteristics may give very misleading indications of the actual inputs to any given student.

Frequently, educational production functions are interpreted as if the included variables are conceptually and accurately measured, when in fact this is not the case. However, the severity of such problems differs significantly across studies and clearly explains part of the apparent inconsistency in findings. Moreover, within most studies, measurement errors are probably most important in the case of school inputs, leading in general to underestimates of the importance of school inputs.²⁴

23 At the elementary school level, current measures of peers may not in general be too bad, i.e., may not have particularly large measurement errors. However, the problem worsens significantly at later grades where it is common to collect students from a variety of schools at a given junior high or high school. Nevertheless, for some analyses such as the effects of racial composition of peers, the situation is more complicated, even at the elementary school level. Commonly, schools observed to be well integrated (those with, say, 20 to 60 percent black students) at any point in time are actually quite unstable—they are going through a transition period. This implies that the current racial composition, particularly in the midranges, may not accurately reflect historic racial composition. Similarly, for schools where the racial composition has recently changed as with redistricting or busing, the current situation may be a poor indicator of the historic situation.

24 Background characteristics are generally measured by a variety of socioeconomic variables. Conceptually, the variables should measure the direct learning provided in the home along

Typically, the school inputs used in empirical analyses include objective measures of teacher characteristics and schools such as education levels, teacher experience, and age of school buildings. Some also include more detailed aspects of teachers such as undergraduate majors or teacher verbal ability, which can be interpreted as simply attempting to measure homogeneous "quantities" of inputs. While these fit clearly in the conceptual framework, consideration of another set of perhaps important variables (such as measures of the organizational context of schools and the educational process itself) introduces a set of conceptual problems with the production function terminology and framework. Typically (outside of education), measures of organization and process are seen as irrelevant in estimation. Production functions are interpreted as the relationship between inputs and outputs *mutatis mutandis*. Information about production possibilities is essentially viewed as being publicly available in the form of scientific and engineering knowledge, and production processes are reproducible through blueprints and machinery. The possibility of dynamic choices in process made by the actors in production is not considered, and the choice of "best" process is assumed automatically made after selection of inputs. While the appropriateness of this framework is open to question in a wide number of instances, it is particularly questionable in the case of education.²⁵

In the education context, there are two separable classes of issues. First, there are observable "macro" organizational and process characteristics of the school (such as class organization, curricula, departmentalization, length of the school day, etc.) which represent clearly defined and reproducible educational practices. Second, there are aspects of the process which are difficult to disentangle from the characteristics of individual teachers (such as classroom management, methods of presenting abstract ideas, communication skills, etc.).

with attitudes, etc. Past research suggests that the learning environment in the home is highly correlated with SES. Empirically, the models do not appear to be sensitive to the precise measure of SES used. For policy considerations, however, it is important to note that the SES measures are only proxies for some more fundamental characteristics, and it is unlikely that changing the measured characteristic (say, current income) will have much of a short-run effect on achievement. At times we might be interested in determining whether there are any immediate effects of changing the measured SES of a family (say, through a negative income tax). However, it should be recognized that this is a reduced-form relationship (where there is another structural equation which relates learning environment in the home to current attributes) and that policies aimed at SES simply might alter the structural relationships.

- 25 The importance of "process" choice is apparent in any activities which involve individual "skill"; e.g., the differences between chefs is probably not just a difference in cookbooks, or blueprints. Organizational issues have been treated to some extent such as models of learning-by-doing, but in general have not been very well developed. See Nelson and Winter [11] for a more general critique of the shortcomings of the engineering view of production functions.

The first set of factors can readily be accommodated in the conceptual framework (although the actual empirical implementation may be more difficult). While decision-makers may not accurately perceive the impact of various macro organization and process choices and thus may not make the best choices, production functions can be estimated conditional upon these factors. In fact, there has been some, although not extensive, investigation along these lines.²⁶

However, the second type of process effect creates more serious problems—both for the application of the general conceptual model and for the interpretation of any estimated effects. Many educational decisions are “micro” ones made by the actors themselves—mainly teachers. These are both difficult to observe and measure and, quite possibly, not easily reproduced. As a shorthand description, these factors will be referred to simply as “skill” differences. Once the possibility of skill differences, or embodied process in individuals, is introduced, the language—if not the conceptual framework—of production functions begins to fail. It is even difficult to define just what “maximum possible output” might mean since it is difficult to specify what the “homogeneous” inputs are.

There is some indication that these latter individual differences are quite important. The explanation of the apparent insignificance of macro process variables in Armor and others [6] is the great variation in implementation of overall process decisions at the classroom level. This is also supported by detailed analysis of the implementation of innovative techniques at the classroom level (see Berman and McLaughlin [15]). Finally, more direct analysis indicates that roughly only half of total teacher performance can be explained by measured teacher and classroom attributes.²⁷

26 For example, Armor and others [6] test a variety of macro organization and process variables including techniques of reading instruction, time spent on reading, team-teaching, open classrooms, and variety of materials used, but found no significant impacts on achievement. In empirical work, organizational forms or process which represent simple, well-defined choices (such as the use of a given standard curriculum) are easily included. However, more complicated or less well-defined factors (such as departmentalization which depends not only upon the organization, but also the particular teachers) present more formidable problems that are related to the second category of process effects (below).

There have also been a large number of direct investigations of alternative processes or organizational forms, generally following experimental approaches and thus having a more narrow focus. See, for example, Jamison and others [74], Carpenter and Hall [31], Garfinkel and Gramlich [49], Gramlich and Koshel [52], Cicirelli and others [35], Armor [5], Barnow and Cain [12], Kiesling [84], Fox [48], and Rivlin and Timpone [121]. These analyses have fairly uniformly shown few achievement effects.

27 This evidence comes from an analysis which first estimates the “value added” of individual teachers through individual teacher dummy variables (Hanushek [61]) and then attempts to explain these estimated differences by measured teacher and classroom variables.

Recognition of skill differences has implications for discussions of "efficiency in production" (discussed below). It also alters our interpretation of teacher and school inputs. It is still reasonable to consider the impact of measured attributes of teachers, since many school decisions such as hiring and salary are based upon a set of these characteristics. However, the estimated impact of these measured attributes, following the above discussion, indicates the ability either to predict or to develop more skilled teachers. For example, the almost universal finding that more education of teachers has no impact on achievement can be interpreted as indicating that teacher training institutions do not, on average, change the skills of teachers. Or, alternatively, the frequent finding that class size doesn't affect achievement may arise from complicated (and unobserved) interactions with teacher process choices; therefore, while it is possible that smaller classes could be beneficial in specific circumstances, it is also true that, in the context of typical school and teacher operations, there is no apparent gain.

One implication of this discussion is that more effort should be devoted to understanding and measuring both the macro and micro organization and process characteristics of schools. This represents a distinct break from the tradition of production function analysis. There is no presumption that schools systematically choose the best process given the inputs; thus, estimates of education "technology" must be made conditional upon the chosen macro organization and process characteristics. At the individual teacher level, the estimated impact of teacher characteristics can be thought of as reduced from coefficients which include both direct effects (say, of teacher experience) and indirect effects through systematic choice of micro process.

D. Efficiency in Production

One important issue is whether or not schools are efficient in production. This has important policy implications since inefficiency indicates the possibility of increasing school outputs with no additional inputs. However, there is a prior statistical and interpretive issue: Since estimation is based upon the observed behavior of schools, the estimated relationships may not trace out the production frontier if schools are not producing the maximum output for given inputs. In such cases, the relationships will describe average behavior which may not be particularly useful in predicting how changes in inputs would affect outputs. Past discussions of efficiency have nevertheless been confused because both the concepts of efficiency being applied and the appropriate ones in this case have not been clear.

Traditionally, two concepts of efficiency are considered. Economic efficiency refers to the correct choice of input mix given the prices of inputs (and the production function). Technical efficiency refers to operating on

the production frontier, that is, maximizing output for a given set of inputs. Past efficiency discussions have blurred these two concepts and, more importantly, have neglected consideration of how expanding the usual concept of production functions to recognize both macro organization and process choice and skill differences of inputs affects efficiency discussions.

Two arguments have been used to support the assertion that schools are technically inefficient.²⁸ First, educational decision-makers are apparently not guided by incentives to maximize profits or to conserve on costs. Second, they might not understand the production process and therefore can't be expected to be on the production frontier. The first argument, while raising the possibility of economic inefficiency, does not necessarily imply being off the production frontier unless resources are also wantonly squandered.²⁹ The focus of the second argument generally appears related to the importance of macro organizational and process choices. The relevance of these can be analyzed, and, importantly, their presence does not significantly alter the interpretation of empirical analyses as production functions.³⁰ Direct analyses of these factors, while not completely conclusive, do not indicate their overwhelming importance (see above).

The possibility of skill (or "embodied process") differences among inputs to schooling introduces a new dimension to the efficiency discussion. The standard conceptual framework indicates that, if two production processes are using the same inputs, any systematic difference in outputs reflects inefficiency. However, the concept of skill differences simply recognizes that individuals with the same measured characteristics make a series of important production decisions (reflected in behavior, process choices, etc.) that are difficult to identify, measure, and model. Therefore, it is not surprising that the same measured inputs yield variations in output, but at the same time it is difficult to label such observed variation as efficiency differences.

Introduction of skill differences does not, however, eliminate the usefulness of a general production framework. For many purposes, the

28 In most production function estimation outside of education, it is assumed that profit motivation dictates efficiency. An exception is Leibenstein [88] who argues that production inefficiency is more common than generally assumed.

29 In fact, if all schools were economically efficient and input prices were the same everywhere, it would not be possible to estimate production functions. The production function is, in education and in other areas, identified either by input price variations or by economic inefficiency.

30 As suggested above, production functions can be estimated conditional upon macro organization and process choice. Comparison of the "best" technology with alternative ones then provides estimates of this source of inefficiency. Clearly, if these factors are important and correlated with observed input usage, estimation not considering them would be misleading (i.e., would describe average relationships).

desired information is what aspects of teaching can be replicated (or predicted) in different situations. Most research has concentrated upon systematic, measurable characteristics—the reduced-form models of teacher effects—and these estimates do indicate what can be replicated in the absence of shifts in the underlying structure.

Some research has in fact estimated the total effects of individual teachers without regard to actual measurement of underlying attributes and confirms that important dimensions of teacher quality are not captured by measured teacher attributes.³¹ An important sidelight of such investigations is that decision-makers might be able to identify underlying skill differences among teachers with fair accuracy. Murnane [106] found that principals' evaluations of teachers were highly correlated with estimates of total effectiveness. For many purposes, this is almost as good as the ability to identify differences *ex ante*.³²

Finally, concern about technical inefficiency has led to some, basically nonstatistical, estimation of the production frontier.³³ Besides assuming accurate measures of both inputs and outputs, this analysis appears internally inconsistent: it is motivated by the perceived uncertainty about the production process, yet assumes that the researcher knows and measures all of the inputs to the production process. Further, the possibility of nonreproducible skill differences is totally neglected.

31 Several studies have been conducted without resort to just measured teacher characteristics such as experience, education, etc., but have used general covariance analysis schemes to estimate the "total" effect of teachers—in terms of both measured and unmeasured characteristics (see Hanushek [61], Murnane [106], or Armor and others [6]). Similar analysis (Hanushek [63]) for "skill" differences among principals, while somewhat limited, show no differences.

32 This indicates that accountability of individual teachers is not impossible. There is some concern, however, about just how this information might be used for educational policy. The principal observations used by Murnane took place in a situation where they were not actually used for any decisions. (For interpretation of these results, it should be pointed out that the evaluations were correlated with the value added of teachers and not simply the overall level of student achievement. Thus, it is not the case that principals' evaluations just reflected observed total performance of students.)

33 One proposed technique is linear programming analysis (Aigner and Chu [1], Carlson [30], Levin [91]). This technique effectively locates the "most productive" schools and places the production plane through just these observations. In fact, the estimation is based upon a small set of observations (equal to the dimensionality of the inputs).

A related inquiry by Klitgaard and Hall [85] attempts to identify "unusually effective" schools, i.e., schools which perform significantly better than would be expected on the basis of SES composition of students and community characteristics. Such schools are identified on the basis of residuals from achievement regressed on SES of students. However, if school attributes are important and correlated with SES, this approach can be very misleading—particularly when important "school variables" are subsequently identified. The residuals simply cannot be interpreted in this manner because they will be biased and inconsistent.

E. Miscellaneous Issues and Nonissues

A number of more detailed criticisms of estimated educational production functions have also surfaced. This section briefly discusses the major ones (functional form, level of aggregation, selection effects, multicollinearity among inputs, and general statistical methodology), along with their effect on the interpretation and conduct of production analyses.

1. *Functional Form.* Educational production functions have been estimated in a variety of forms, although most frequently variants of linear or logarithmic models.³⁴ Conceptually, there is little guidance about functional form.³⁵ Empirically, the issue of functional form appears to be a second order problem, since distinguishing among alternative functional forms is often impossible. The point is a simple one: within a limited range of variation, many functional forms look very similar. An important corollary, however, is that over wider ranges of variation, different functional forms may yield very different results, implying that predictions based upon changes that are far from current observations may be perilous. For example, while variations in class size, within the limited ranges observed, have little apparent effect on achievement, this is not necessarily the case for radically different class sizes.

2. *Level of Aggregation.* While the conceptual model is at the individual student level, much analysis—relying upon data collected for other purposes—is actually conducted at a more aggregate level, say, the school or district level. The effects on the estimates of such aggregation depend crucially upon the nature of educational relationships.

In the simplest case, when the production process is approximately linear in the same parameters for all students, OLS estimates on the aggregate data will be unbiased, although probably less precise than if individual data were available.³⁶ In more complicated situations, aggregation

34 The variety is actually fairly large. Various authors have considered various stratifications (such as by race or socioeconomic background); e.g., Hanushek [61], Coleman and others [37], Smith [128]. Others have conducted general covariance analyses which allow unconstrained functional forms in terms of underlying descriptors of teachers (Hanushek [61], Murnane [106]). Finally, a variety of interactions among variables have been introduced (Winkler [149], Summers and Wolfe [133, 134]).

35 Linear models imply independence of the various inputs and constant marginal products, while logarithmic models allow declining marginal products but constrain the form of interactions of variables. Much production function analysis outside of education has centered on the properties and usefulness of alternative functional forms. See, for example, the work on Cobb-Douglas forms in Hildebrand and Liu [71]; constant elasticity of substitution models in Arrow and others [8]; transcendental log functions in Christensen, Jorgenson, and Lau [34]; and generalized production functions in Hanoch [59]. However, as explained previously, these analyses are difficult to translate for education.

36 More generally, even with a distribution of parameters across students, OLS estimates of the mean parameters are unbiased as long as the parameters are uncorrelated with the

has less innocuous effects. For example, if two groups of students—say, blacks and whites—have different production relationships where the differences are not easily parameterized, estimation with aggregate data yields “average” coefficients which depend upon the weighting of the two groups in the sampled observations and which are difficult to interpret. (See Stodolsky and Lesser [132] on the possibility of racial differences.)

Nevertheless, probably the most serious “aggregation” problem is really one of errors of measurement. The researcher frequently has individual data about students (say, achievement and family background), but only aggregate data about schools. The temptation is to use all available data by mixing individual characteristics with aggregate school data. However, the school factors relevant to any individual may differ significantly from the average. (Consider, for example, the situation in a large comprehensive high school.) Here, aggregation generally helps; the errors in measurement for a model of average achievement and average characteristics are almost certainly less than with individual achievement and average school characteristics.³⁷

Aggregation as an errors-in-variables problem may be quite pervasive. Even with data on individual classrooms, the internal allocation of time and attention of students implies that each student might receive different inputs (see, for example, Garner [50], Wiley and Harnischfeger [148], and Karweit [79]). We don’t have a good understanding of the importance of such variations. Analysis of classroom composition effects can be interpreted as attempts to model this, but the results about the importance of peer compositions are mixed.³⁸ As discussed previously, the accuracy of measurement of inputs is an extremely important topic, and probably much more important than just aggregation.

3. *Selection Effects and Causation.* For policy purposes, information about causal relationships between school factors and achievement is needed. However, such information (about the direction of causation) cannot come directly from the observed data and correlations, but must be introduced from a priori information about the structure of the overall

exogenous variables and have the same mean across students; see Swamy [135] for discussion of such “random” coefficient models. With both random coefficients and simple aggregation, the efficiency of estimation can generally be improved with techniques other than OLS (see Hanushek and Jackson [65]). Such results do not, however, hold when there are important nonlinearities in the production process.

37 This errors-in-variables argument is a major criticism of the original Coleman work (see Hanushek and Kain [66] and Hanushek [61]). It is particularly damaging in the analysis of variance framework of Coleman and others [37].

38 Compare Hanushek [61] with Henderson and others [70] for peer estimates at the classroom level. See also Murnane [106].

model. The primary concern in the production function setting is the effects of teacher selection and assignment mechanisms.

Consider the simple case of an observed positive relationship between teacher experience and student achievement (holding other factors constant). Depending upon the mechanism by which teachers are assigned to schools, this need not imply that increasing average experience levels in a school will increase achievement (i.e., that there is a causal relationship running from experience to achievement). If, for example, more senior teachers were allowed to choose their schools and teachers had a preference for teaching higher achieving students, then achievement would, at least in part, "cause" experience; and a policy change that increased experience would not yield the (full) effect on achievement expected from the estimated relationship. Other, and more subtle, selection effects might also occur; more educated or more intelligent teachers may, through their own efforts or the direct assignments of principals, be placed in "faster" classes.

The situation is really another case of simultaneous equation bias. The importance of these effects depends upon the importance of achievement in determining assignments of different types of teachers, and there has been little direct analysis of this. The appropriate solution is estimation of the simultaneous system (see fn. 12), but this has not been done.

Greenberg and McCall [53] analyzed a single urban school system in the early 1970s and concluded that race and socioeconomic background of students were systematically related to the selection and transfer of teachers with different education and experience levels. However, Murnane [107] suggests, from analysis of a different school system, that declining enrollments and the subsequent surplus of teachers have led to a much greater reliance on institutional rules and much less on individual teacher preferences (which was the hypothesized mechanism in Greenberg and McCall [53]).

Nevertheless, the potential problems arise from achievement affecting selection, and not from family background, race, or other factors that are included on the right-hand side of the estimated model affecting selection. In the latter instance (which would be a recursive structure), even though some correlation among the right-hand side variables may be induced by this mechanism, there are generally not serious problems; without other such selection effects, the estimated relationships with achievement can plausibly be interpreted as causal relationships. Clearly the severity of the problem is related to the structure of the model estimated and in many instances is only serious in the presence of fairly subtle selection mechanisms (particularly in a "value-added" specification).

4. *Multicollinearity.* Since discussion of multicollinearity in educational research by Bowles and Levin [24], it is taken as an almost ever-present, but lamentable, fact of life in most estimation. In fact, it is the first item

discussed under "Analytical Problems" in the review of production function research by Averch and others [11]. Fortunately or unfortunately, multicollinearity does not appear to be the villain it has been made out to be, although it may partially explain some of the apparent inconsistencies in existing research.

The statistical story is that disentangling the separate effects of exogenous variables which are very highly intercorrelated can be difficult. Further, in the usual case of positive intercorrelations, the parameter estimates themselves will tend to be negatively correlated so that quite commonly a coefficient has the "wrong sign" because of the correlations of variables.

Nevertheless, the importance of multicollinearity is probably overrated. All correlations of exogenous variables do not have serious consequences, and all low *t*-statistics and wrong signs are not the result of multicollinearity. Right-hand side variables are often called independent variables, but this does not imply that they cannot be correlated. In fact, multiple regression analysis is used *because* there are correlations among the "independent" variables.

The importance of multicollinearity depends crucially upon the statistical methods used.³⁹ Further, the diagnosis of problems is sometimes difficult.⁴⁰ In micro or individual level data, such as those frequently used, the real estimation problems caused by multicollinearity are almost certainly not as severe as citations would indicate. This problem could, however, explain some of the variation in findings across studies since, in the absence of an agreed-upon theory of what variables should be included in the models and how they should be measured, researchers frequently determine model specification on the basis of coefficient significance tests. Variations in the sample intercorrelations will yield variations in model specifications under such criteria.

39 Bowles and Levin [24] were correct in the importance of multicollinearity for the Coleman analysis. However, the important feature in that criticism, that is often overlooked or misunderstood, is the interaction between the choice of statistical analysis (analysis of variance) and the correlation of exogenous variables. In the analysis of variance framework, *any* correlation causes trouble. See also Hanushek and Kain [66] and the next section of this paper.

40 Multivariate regression analysis is designed to take into account correlations among the exogenous variables. If the exogenous variables are uncorrelated, bivariate regression (or simple correlations) will suffice. However, with very high levels of correlation among the independent variables, the coefficient estimates become imprecise; in the extreme, with perfect linear relationships among exogenous variables, estimation of the independent effects of the variables is simply impossible. Diagnosis is often difficult because multicollinearity often causes high estimated variances of the coefficient, but such high variance can also result from including variables which are unimportant in determining achievement. For further discussion and diagnostic aids, see Hanushek and Jackson [65].

5. *Statistical Methods.* Throughout the previous discussion, the focus has been on the estimation of the parameters of the production process. An alternative focus, and a close statistical relative, is contained in analysis of variance (see Coleman and others [37]). In this methodology, the observed achievement variance is decomposed into that arising from different sources. Suffice it to say, this methodology is often inappropriate for the questions under consideration (see Cain and Watts [28]), and simply raises further, added interpretive questions with no apparent gain.⁴¹

IV. AN ASSESSMENT OF WHAT WE KNOW AND WHAT WE SHOULD DO

The discussion to this point has indicated a wide range of problems—from conceptual problems to technical and esoteric interpretive issues. However, the overall analytical power of the production function framework, which integrates observations about various aspects of schools, should not be lost. Clearly, more detailed theoretical and empirical analyses focusing on specific aspects of the production process (such as the mechanisms of peer influences or the organizational and decision-making framework of schools) have been conducted outside of the context of production function analysis. However, they generally suffer from one of two problems: Either they concentrate exclusively upon a given attribute of schools or the learning process, or they consider the relationship between a particular attribute and student outcomes to the exclusion of other attributes that simultaneously affect outcomes. While these studies are useful in clarifying the important attributes of schools and in describing what goes on in and around schools, they are considerably less useful in considering alternative policies with respect to schooling. The strength of the production function studies lies in their policy relevance through investigation of the independent influences of various factors—student characteristics, teacher and school inputs, and other environmental attributes—on performance of the schooling system.

Further, this discussion should not be interpreted as implying that we have not learned anything from past research. In fact, there are some startlingly consistent findings that are quite robust to some of the problems mentioned, and many apparent inconsistencies are significantly reduced

41 The results of these techniques are sample specific; that is, they depend importantly upon the observed sample variations in the dependent and independent variables. Further, some of the variation, maybe a significant portion of it, can be accounted for jointly by the various independent variables (see Hanushek and Kain [66]), which leaves the choice of arbitrarily allocating the joint variation (as in Coleman and others [37]) or simply identifying its importance (as in Mayeske and others [99]). In neither case can one indicate the expected effect of changing given inputs.

when the most obvious specification, estimation, and measurement errors are taken into account. Not only are these problems not uniformly distributed across studies, but also it is often possible to determine the direction, if not the plausible magnitude, of many such biases.

In terms of consistent findings, differences in family socioeconomic background without question lead to significant achievement differences. Socioeconomic status is interpreted as a proxy for quality of the home learning environment. How much arises from factors malleable in the short run (current income and consumption, physical surroundings, current attitudes, etc.) and how much arises in the longer run, less malleable attributes (such as parental education, patterns of family rearing, etc.) is unknown, although longer run attributes are probably more important.⁴²

Second, there is conclusive evidence that differences among schools and teachers are important in achievement. Schools simply do not have homogeneous impacts on students. Yet, the identification and measurement of specific teacher or school attributes which are important is much less certain. The variation across studies in specific characteristics that appear important in part reflects incomparabilities in underlying samples and data. For example, measures of "general intelligence" of teachers appear consistently important when considered (see Hanushek [61]), but are most frequently unavailable. Part of the variation undoubtedly also reflects teacher "skill" differences that are difficult to identify, measure, and model. Nevertheless, there is some evidence that school officials can identify more productive teachers. While the inability to disentangle the attributes of effective teachers indicates difficulty in selecting "good" teachers *ex ante* or in improving teacher productivity, the fact that good teachers can be identified *ex post* indicates that schools can be improved by appropriate promotion and allocation decisions.

Third, there is quite conclusive evidence that schools are economically inefficient; that is, they do not employ the best mixes of inputs, given input prices and their apparent effectiveness.⁴³ The possibility of inefficiency in

42 Analysis of this issue, which deserves added attention, requires longitudinal data on individuals, preferably where current family characteristics change significantly. (Data generated from the negative income tax experiments in Gary, Seattle, and Denver, for example, seem appropriate). Further, it may be even more useful to actually model behavior within households. Nevertheless, simply because a current background measure, say, family income, appears important, one cannot conclude that changing this will increase achievement in the short run since it may proxy other attributes (that aren't changed) and may be contaminated by individuals' ability effects (see above).

43 The evidence on economic efficiency comes from two, almost universal, findings of no consistent or significant relationship: (1) between achievement and expenditures per pupil (either instructional expenditure or total expenditures); and (2) between achievement and specific purchased inputs (teacher experience, teacher education levels, class size, and administrative/supervisory expenditures). Teacher experience appears somewhat produc-

school choices of organization and process, while less conclusive, does not however appear overwhelming.⁴⁴

Fourth, there are significant differences in production functions by race and, perhaps, family background; that is, school resources interact importantly with the background characteristics of individuals.⁴⁵ On the research side, this implies exercising considerable care in modeling efforts, particularly when forced to use aggregate data. On the policy side, it indicates that more attention should be given to the internal allocation of resources. If inputs were equally effective for all students, shuffling teachers (without changing the pool) is a zero-sum game—winners balance against losers; this is not the case when input effectiveness varies across students.

In addition to these significant substantive results, there are, however, a number of significant gaps in our knowledge. Several of these have come out in the previous discussion and require only summarizing here.

The primary gap in understanding—at both a conceptual and empirical level—is an inadequate picture of the relationship between school quality and subsequent performance and therefore of how to measure school quality. The previous findings on the operations of the school system relate almost exclusively to test-score measures of achievement, even though validity of this measure is quite uncertain. It is difficult to overemphasize the importance of pursuing this line of inquiry.⁴⁶

Beyond this, there are also a series of uncertainties which are amenable to research. For example, the influence of peer compositions—which is central to such important questions as integration and ability-tracking policies—remains murky.⁴⁷ Also, the dynamics of the educational process

tive, but clearly less productive at the margin than its cost based upon typical salary schedules. The others do not appear to show any positive relationships, let alone relationships strong enough to justify their costs.

44 Relatedly, little information about the decision process which dictates organizational and process determination is available. While not essential for understanding the production process per se, this might be very relevant for developing appropriate strategies to institute change. Available work on introduction and implementation of "innovations" (really process changes) includes Berman and McLaughlin [15] and Silkman [127].

45 Previous work indicates that stratification by race or SES or inclusion of interaction terms among school resources and backgrounds consistently shows significant achievement relationship differences: for example, Hanushek [61] or Summers and Wolfe [134].

46 While data availability is a clear problem, more can be done along these lines even with existing data. Considerable individual data with information about qualitative aspects of schooling are becoming available, but have not been fully exploited for linkages with education analyses. For example, Sewell and Hauser [126] have data about specific schools attended and lifetime outcomes, as does part of the Survey of Income and Education. The Michigan Panel Study of Income Dynamics and the National Longitudinal Surveys have a variety of test measures along with outcome data. Each could potentially provide more information about school quality and subsequent outcomes.

47 In many studies, data limitations (either because of aggregation problems or incomplete

are imperfectly understood. Most studies have been cross-sectional and rely upon specialized samples with little information about the impacts of resources for different age-school year cohorts. Therefore, given other incommensurates, it has been very difficult to analyze such issues as the differential impact of early school programs or varying achievement patterns.⁴⁸ As a final example, there has been little investigation of the stability of teacher skills, or individual teacher effects, over time.

Many of the analytical problems are ones where the conceptual problems are minimal, but data problems are severe.⁴⁹ In particular, the current methodology is probably quite adequate for many further analyses of elementary education where, among other things, there are fewer measurement problems and generally simpler organizational structures. There simply seem to be large, immediate payoffs to collecting new data for a variety of school situations where the most significant measurement questions indicated above are avoided and where longitudinal information can be obtained.⁵⁰

Another set of problems—ones where conceptual inadequacies appear paramount—is nevertheless very significant. They have occurred in the previous discussion and include measures of alternative outputs, investigation of decision processes with regard to alternative output mixes, identification and measurement of process and organizational variables for both schools and classrooms, and expansion of our models for more complicated realities such as high schools.

V. CONCLUDING REMARKS

While the primary motivation for analyzing educational production relationships is derived from the sector's important resource usage, from its

information) preclude analysis of peer influences. The studies directly considering peer influences (Hanushek [61] and Henderson and others [70]) yield conflicting results. The research into racial composition has generally been flawed by limited historical information.

48 Several studies have included measures of preschool programs (e.g., Hanushek [61], Ritzén and Winkler [120]). These generally show more impact than is found in direct analyses such as that of Headstart (Cicirelli and others [35]).

49 One clear problem has been the use of specialized samples which relate to limited situations and which (because generally collected for other purposes) are not always appropriate. Historically, however, there seems to be a bias against data collection. Considerable money was spent on reanalyzing the data from the Coleman Report, even though it was well recognized that there were many, near-fatal flaws in those data, while none was spent on simply gathering better data.

50 A panel study design could, in addition to allowing dynamic analyses, be used to minimize remaining problems such as missing data on individual abilities; see Boardman and Murnane [19] or Hanushek and Jackson [65].

importance as a policy instrument, and from concerns about efficiency given the lack of market incentives, such analyses also have direct linkages to other research and policies. For example, the entire focus of school finance system discussions has been the distribution of expenditures per student under alternative financing mechanisms, even though the apparent levels of economic inefficiency in education indicate that this has little to do with the distribution of educational services. Similarly, many studies (found in public finance or urban economics) control for the quality of governmental services by including school expenditures per student with little consideration of what this is actually measuring. While part of school integration discussions relate to the effect of racial composition on achievement, ambiguities in this work have led to considerable confusion in this area (see Clune [36]). As a final example, even though years of schooling is a clearly inadequate measure of individual skill and ability differences, most contemporary labor economics and sociological research deals at only this level.

While other examples are easy to compile, the message should be clear: Understanding the educational sector has important ramifications for understanding many other areas, but the treatment of education has, for the most part, been quite superficial.

REFERENCES

1. Dennis J. Aigner and S. F. Chu. "On Estimating the Industry Production Function." *American Economic Review* 58 (September 1968).
2. John S. Akin and Irwin Garfinkel. "School Expenditures and the Economic Returns to Schooling." *Journal of Human Resources* 12 (Fall 1977): 460-81.
3. Karl L. Alexander and Edward L. McDill. "Selection and Allocation Within Schools: Some Causes and Consequences of Curriculum Placement." *American Sociological Review* (December 1976).
4. Duane Alwin. "College Effects on Educational and Occupational Attainments." *American Sociological Review* (April 1974).
5. David Armor. "The Evidence on Busing." *Public Interest* (Summer 1972).
6. David Armor and others. *Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools*. R-2007-LAUD. Santa Monica, Calif.: Rand Corporation, 1976.
7. Kenneth J. Arrow. "Higher Education as a Filter." *Journal of Public Economics* 2 (1973).
8. Kenneth J. Arrow and others. "Capital-Labor Substitution and Economic Efficiency." *Review of Economics and Statistics* (August 1961).
9. Orley Ashenfelter. "Estimating the Effects of Training on Earnings." *Review of Economics and Statistics* (February 1978).
10. Orley Ashenfelter and Joseph D. Mooney. "Graduate Education, Ability, and Earnings." *Review of Economics and Statistics* (February 1968): 78-86.

11. Harvey A. Averch and others. *How Effective Is Schooling?* Englewood Cliffs, N.J.: Educational Technology Publications, 1974.
12. Burt S. Barnow and Glen G. Cain. "A Reanalysis of the Effect of Head Start on Cognitive Development: Methodology and Empirical Findings." *Journal of Human Resources* 12 (Spring 1977): 177-97.
13. Yoram Ben-Porath. "The Production of Human Capital Over Time." In *Education, Income, and Human Capital*, ed. W. Lee Hansen. New York: National Bureau of Economic Research, 1970.
14. Ivar E. Berg. *Education and Jobs: The Great Training Robbery*. New York: Praeger Publishers, 1970.
15. Paul Berman and Milbrey W. McLaughlin. *Federal Programs Supporting Educational Change, Vol. IV: The Findings in Review*. R-1589/4-HEW. Santa Monica, Calif.: Rand Corporation, 1975.
16. Matthew Black. "The Determinants and Implications of Job Search and Quit Activity." Ph.D. dissertation, University of Michigan, 1976.
17. Peter Blau and Otis D. Duncan. *The American Occupational Structure*. New York: John Wiley & Sons, 1967.
18. Anthony E. Boardman, Otto A. Davis, and Peggy R. Sanday. "A Simultaneous Equations Model of the Educational Process." Pittsburgh: School of Urban and Public Affairs, Carnegie-Mellon University, 1973, mimeo.
19. Anthony E. Boardman and Richard J. Murnane. "The Use of Panel Data in Education Research." Fels Discussion Paper No. 110. Philadelphia: University of Pennsylvania, 1977.
20. Raymond Boudon. *Education, Opportunity, and Social Inequality*. New York: John Wiley & Sons, 1974.
21. Samuel Bowles. "Towards an Education Production Function." In *Education, Income, and Human Capital*, ed. W. Lee Hansen. New York: National Bureau of Economic Research, 1970.
22. Samuel Bowles and Herbert Gintis. *Schooling in Capitalist America*. New York: Basic Books, 1976.
23. Samuel Bowles and Henry M. Levin. "The Determinants of Scholastic Achievement—A Critical Appraisal of Some Recent Evidence." *Journal of Human Resources* 3 (Winter 1968): 3-24.
24. ———. "More on Multicollinearity and the Effectiveness of Schools." *Journal of Human Resources* 3 (Summer 1968): 393-400.
25. Byron W. Brown and Daniel H. Saks. "The Production and Distribution of Cognitive Skills Within Schools." *Journal of Political Economy* 83 (April 1975).
26. Jesse Burkhead, with Thomas G. Fox and John W. Holland. *Input and Output in Large-City Schools*. Syracuse, N.Y.: Syracuse University Press, 1967.
27. Glen G. Cain and Robinson G. Hollister. "The Methodology for Evaluating Social Action Programs." In *An Anthology of Labor Economics: Readings and Commentary*, eds. Ray Marshall and Richard Perlman. New York: John Wiley & Sons, 1972.
28. Glen G. Cain and Harold W. Watts. "Problems in Making Policy Inferences from the Coleman Report." *American Sociological Review* (April 1970).
29. Angus Campbell and others. *The American Voter*. Chicago: University of Chicago Press, 1976.

30. Daryl Carlson. "Examining Efficient Joint Production Processes." In *Measuring and Increasing Academic Productivity*, ed. Robert A. Wallhaus. New Directions for Institutional Research No. 8, Winter 1975.
31. Polly Carpenter and George R. Hall. *Case Studies in Educational Performance Contracting: Conclusions and Implications*. R-900/1-HEW. Santa Monica, Calif.: Rand Corporation, 1971.
32. Central Advisory Council for Education (England). *Children and Their Primary Schools*. London: Her Majesty's Stationery Office, 1967.
33. Barry R. Chiswick. "Schools and Earnings of Low Achievers: Comment." *American Economic Review* 62 (September 1972): 752-54.
34. Laurits R. Christensen, Dale W. Jorgenson, and Laurence J. Lau. "Transcendental Logarithmic Production Frontiers." *Review of Economics and Statistics* (February 1973).
35. Victor G. Cicirelli and others. *The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development*. Westinghouse Learning Corporation, 1969.
36. William H. Clune. "Law and Economics in *Hobson v. Hansen*—An Introductory Note." *Journal of Human Resources* 7 (Summer 1972): 275-82.
37. James S. Coleman and others. *Equality of Educational Opportunity*. Washington: U.S. Government Printing Office, 1966.
38. James S. Coleman. "The Evaluation of 'Equality of Educational Opportunity.'" In *On Equality of Educational Opportunity*, eds. F. Mosteller and D. P. Moynihan. New York: Random House, 1972.
39. James S. Coleman and Nancy L. Karweit. *Measures of School Performance*. R-488-RC. Santa Monica, Calif.: Rand Corporation, 1970.
40. Lee J. Cronbach and Lita Furby. "How Should We Measure 'Change'—or Should We?" *Psychology Bulletin* (1970).
41. Edward F. Denison. *Accounting for United States Economic Growth 1929-69*. Washington: Brookings Institution, 1974.
42. Dennis J. Dugan. "Scholastic Achievement: Its Determinants and Effects in the Education Industry." In *Education as an Industry*, eds. Joseph T. Froomkin, Dean T. Jamison, and Roy Radner. Cambridge, Mass.: Ballinger, 1976.
43. Greg J. Duncan. "Nonpecuniary Work Rewards." In *Five Thousand American Families: Patterns of Economic Progress*, Vol. II, ed. James N. Morgan. Ann Arbor: Institute for Social Research, University of Michigan, 1974.
44. Otis D. Duncan, David Featherman, and Beverly Duncan. *Socioeconomic Background and Achievement*. New York: Seminar Press, 1972.
45. Rupert N. Evans and Joel D. Galloway. "Verbal Ability and Socioeconomic Status of 9th and 12th Grade College Preparatory, General, and Vocational Students." *Journal of Human Resources* 8 (Winter 1973): 24-36.
46. Franklin M. Fisher. *The Identification Problem in Econometrics*. New York: McGraw-Hill, 1966.
47. Manley Fleischmann and others. *Report of the New York State Commission on the Quality, Cost, and Financing of Elementary and Secondary Education*. 1978.

48. David J. Fox. "Expansion of the More Effective School Program." New York: Center for Urban Education, 1967, mimeo.
49. Irwin Garfinkel and Edward M. Gramlich. "A Statistical Analysis of the OEO Experiment in Educational Performance Contracting." *Journal of Human Resources* 8 (Summer 1973): 275-305.
50. William T. Garner. "The Identification of an Education Production Function by Experimental Means." Ph.D. dissertation, University of Chicago, 1973.
51. Herbert J. Gintis. "Education, Technology, and the Characteristics of Worker Productivity." *American Economic Review* 61 (May 1971).
52. Edward M. Gramlich and Patricia P. Koshel. *Educational Performance Contracting*. Washington: Brookings Institution, 1975.
53. David Greenberg and John McCall. "Teacher Mobility and Allocation." *Journal of Human Resources* 9 (Fall 1974): 480-502.
54. Zvi Griliches. "Production Functions in Manufacturing: Some Preliminary Results." In *The Theory and Empirical Analysis of Production*, ed. Murray Brown. New York: National Bureau of Economic Research, 1967.
55. ———. "Estimating the Returns to Schooling: Some Econometric Problems." *Econometrica* (January 1977).
56. Zvi Griliches and William Mason. "Education, Income, and Ability." *Journal of Political Economy* 80 (May/June 1972, Pt. 2).
57. Michael Grossman. *The Demand for Health: A Theoretical and Empirical Analysis*. National Bureau of Economic Research Occasional Paper 119. New York: Columbia University Press, 1972.
58. James W. Guthrie and others. *Schools and Inequality*. Cambridge, Mass.: MIT Press, 1971.
59. Giora Hanoch. "Production and Demand Models with Direct or Indirect Implicit Additivity." *Econometrica* (May 1975).
60. W. Lee Hansen, Burton Weisbrod, and William J. Scanlon. "Schooling and Earnings of Low Achievers." *American Economic Review* 60 (June 1970).
61. Eric A. Hanushek. *Education and Race*. Lexington, Mass.: D.C. Heath, 1972.
62. ———. "Regional Differences in the Structure of Earnings." *Review of Economics and Statistics* (May 1973).
63. ———. "Comment." In *Education as an Industry*, eds. Joseph T. Froomkin and others. Cambridge, Mass.: Ballinger, 1976.
64. ———. "Ethnic Income Variations: Magnitudes and Explanations." In *American Ethnic Groups*, ed. Thomas Sowell. Washington: Urban Institute, 1978.
65. Eric A. Hanushek and John E. Jackson. *Statistical Methods for Social Scientists*. New York: Academic Press, 1977.
66. Eric A. Hanushek and John F. Kain. "On the Value of 'Equality of Educational Opportunity' as a Guide to Public Policy." In *On Equality of Educational Opportunity*, eds. F. Mosteller and D. P. Moynihan. New York: Random House, 1972.
67. Eric A. Hanushek and John M. Quigley. "Implicit Investment Profiles and Intertemporal Adjustments of Relative Wages." *American Economic Review* 68 (March 1978).

68. Donald P. Hayes and Judith Grether. "The School Year and Vacations: When Do Students Learn?" Washington: Urban Institute, 1969.
69. John Heim and Lewis Perl. *The Educational Production Frontier*. Ithaca, N.Y.: Cornell University, 1974.
70. Vernon Henderson, Peter Mieszkowski, and Yvon Sauvageau. *Peer Group Effects and Educational Production Functions*. Ottawa: Economic Council of Canada, 1976.
71. George H. Hildebrand and T. C. Liu. *Manufacturing Production Functions in the U.S., 1957*. Ithaca, N.Y.: Cornell University Studies in Industrial and Labor Relations, 1965.
72. Russell Hill and Frank Stafford. "Allocation of Time to Preschool Children and Educational Opportunity." *Journal of Human Resources* 9 (Summer 1974): 323-41.
73. Robert P. Inman. "The Fiscal Reform of Metropolitan Schools: Some Simulation Results." *American Economic Review* 68 (March 1978).
74. Dean T. Jamison and others. "Cost and Performance of Computer-Assisted Instruction for Education of Disadvantaged Students." In *Education as an Industry*, eds. Joseph T. Froomkin and others. Cambridge, Mass.: Ballinger, 1976.
75. Christopher S. Jencks. "The Quality of Data Collected by the 'Equality of Educational Opportunity' Survey." In *On Equality of Educational Opportunity*, eds. F. Mosteller and D. P. Moynihan. New York: Random House, 1972.
76. Christopher S. Jencks and others. *Inequality: A Reassessment of the Effects of Family and Schooling in America*. New York: Basic Books, 1972.
77. Christopher S. Jencks and Marsha D. Brown. "Effects of High Schools on Their Students." *Harvard Educational Review* (August 1975).
78. George E. Johnson and Frank P. Stafford. "Social Returns to Quantity and Quality of Schooling." *Journal of Human Resources* 8 (Spring 1973): 139-55.
79. Nancy Karweit. "The Organization of Time in Schools: Time Scales and Learning." Baltimore: Johns Hopkins University, 1978, mimeo.
80. Martin Katzman. *The Political Economy of Urban Schools*. Cambridge, Mass.: Harvard University Press, 1971.
81. Stuart Kerachsky and Charles D. Mallar. "Design of an Evaluation of the Job Corps." Working Paper No. C-12. Princeton, N.J.: Mathematica Policy Research, February 1977.
82. Nicholas Kiefer. "Federally Subsidized Occupational Training and the Employment and Earnings of Male Trainees." *Journal of Econometrics* (forthcoming).
83. Herbert J. Kiesling. "Measuring a Local Government Service: A Study of School Districts in New York State." *Review of Economics and Statistics* (August 1967).
84. ———. "A Study of the Relationship of Instruction Process and Program Organization to the Success of Compensatory Education Projects in California." In *Education as an Industry*, eds. Joseph T. Froomkin and others. Cambridge, Mass.: Ballinger, 1976.
85. Robert E. Klitgaard and George R. Hall. "Are There Unusually Effective Schools?" *Journal of Human Resources* 10 (Winter 1975): 90-106.

86. Richard Layard and George Psacharopoulos. "The Screening Hypothesis and the Returns to Education." *Journal of Political Economy* 82 (September/October 1974).
87. A. James Lee. "The Economics of Compulsory Schooling—A Theoretic and Empirical Investigation." Ph.D. dissertation, University of Wisconsin, 1976.
88. Harvey Leibenstein. "Allocative Efficiency vs. 'X-Efficiency.'" *American Economic Review* 56 (June 1966).
89. Henry M. Levin. "A New Model of School Effectiveness." In *Do Teachers Make a Difference?* U.S. Office of Education. Washington: U.S. Department of Health, Education, and Welfare, 1970.
90. ———, ed. *Community Control of Schools*. Washington: Brookings Institution, 1970.
91. ———. "Concepts of Economic Efficiency and Educational Production." In *Education as an Industry*, eds. Joseph T. Froomkin and others. Cambridge, Mass.: Ballinger, 1976.
92. Frank S. Levy. "The Racial Imbalance Act of Massachusetts." Ph.D. dissertation, Yale University, 1969.
93. Arleen Leibowitz. "Home Investments in Children." *Journal of Political Economy* 82 (March/April 1974, Pt. II).
94. Peter Lindert. "Family Inputs and Inequality Among Children." Discussion Paper No. 218-74. Madison: Institute for Research on Poverty, University of Wisconsin, 1974.
95. Charles R. Link and Edward C. Ratledge. "Social Returns to Quantity and Quality of Education: A Further Statement." *Journal of Human Resources* 10 (Winter 1975): 78–89.
96. ———. "Student Perceptions, I.Q., and Achievement." *Journal of Human Resources* 14 (Winter 1979): 98–111.
97. Larry M. Manheim. "Health, Health Practices, and Socioeconomic Status: The Role of Education." Ph.D. dissertation, University of California, Berkeley, 1975.
98. Stanley Masters and Thomas Ribich. "Schooling and Low Achievers: Comment." *American Economic Review* 62 (September 1972): 755–59.
99. George Mayeske and others. *A Study of the Achievement of Our Nation's Students*. Washington: U.S. Department of Health, Education, and Welfare, 1973.
100. Rebecca Maynard and David L. Crawford. "School Performance." In *Rural Income Maintenance Experiment: Final Report*, Vol. VI, Ch. 12. Madison: Institute for Research on Poverty, University of Wisconsin, 1976.
101. Stephan Michelson. "The Association of Teacher Resources with Children's Characteristics." In *Do Teachers Make a Difference?* U.S. Office of Education. Washington: U.S. Department of Health, Education, and Welfare, 1970.
102. ———. "For the Plaintiffs—Equal School Resource Allocation." *Journal of Human Resources* 7 (Summer 1972): 283–306.
103. Jacob Mincer. "The Distribution of Labor Income: A Survey with Special Reference to the Human Capital Approach." *Journal of Economic Literature* 8 (March 1970).
104. Richard D. Morgenstern. "Direct and Indirect Effects on Earnings of

- Schooling and Socio-Economic Background." *Review of Economics and Statistics* 55 (May 1973): 225-33.
105. Frederick Mosteller and Daniel P. Moynihan, eds. *On Equality of Educational Opportunity*. New York: Random House, 1972.
106. Richard Murnane. *The Impact of School Resources on the Learning of Inner City Children*. Cambridge, Mass.: Ballinger, 1975.
107. ———. "Labor Mobility in a Declining Industry: The Case of Teachers." Paper presented at Econometric Society meetings, August 1978.
108. Richard R. Nelson and Edmund Phelps. "Investment in Humans, Technology, Diffusion, and Economic Growth." *American Economic Review* 56 (May 1966).
109. Richard R. Nelson and Sidney G. Winter. "Neoclassical vs. Evolutionary Theories of Economic Growth: Critique and Prospectus." *Economic Journal* (December 1974).
110. ———. "Factor Price Changes and Factor Substitution in an Evolutionary Model." *Bell Journal of Economics* (Autumn 1975).
111. ———. "Organizational Capability in a Dynamic World: A Behaviorist Reconstruction of the Production Set Idea." New Haven, Conn.: Yale University, October 1976, mimeo.
112. Richard Niemi and Barbara I. Sobieszek. "Political Socialization." *Annual Review of Sociology* (1977).
113. Dave M. O'Neill, Burton Gray, and Stanley Horowitz. "For the Defendants — Educational Equality and Expenditure Equalization Orders." *Journal of Human Resources* 7 (Summer 1972): 307-25.
114. John D. Owen. "The Distribution of Educational Resources in Large American Cities." *Journal of Human Resources* 7 (Winter 1972): 26-38.
115. Lewis J. Perl. "Family Background, Secondary School Expenditure, and Student Ability." *Journal of Human Resources* 8 (Spring 1973): 156-80.
116. President's Commission on School Finance. *Schools, People, and Money*. Washington: U.S. Government Printing Office, 1972.
117. Thomas I. Ribich and James L. Murphy. "The Economic Returns to Increased Educational Spending." *Journal of Human Resources* 10 (Winter 1975): 56-77.
118. John Riew. "Economies of Scale in High School Operations." *Review of Economics and Statistics* (August 1966).
119. John G. Riley. "Competitive Signalling." *Journal of Economic Theory* 10 (1975).
120. Jozef M. Ritzen and Donald R. Winkler. "The Production of Human Capital Over Time." *Review of Economics and Statistics* (November 1977).
121. Alice Rivlin and P. Michael Timpane, eds. *Planned Variation in Education*. Washington: Brookings Institution, 1975.
122. Daniel P. Rogers. "Private Rates of Return to Education in the United States: A Case Study." *Yale Economic Essays* (Spring 1969).
123. Sherwin Rosen. "Human Capital: A Survey of Empirical Research." In *Research in Labor Economics*, Vol. I, ed. Ronald G. Ehrenberg. Greenwich, Conn.: JAI Press, 1977.
124. James E. Rosenbaum. *Making Inequality*. New York: John Wiley & Sons, 1976.
125. Sandra Scarr and Richard A. Weinberg. "The Influence of 'Family Background' on Intellectual Attainment: The Unique Contribution of Adoptive

- Studies for Estimating Environmental Effects." New Haven, Conn.: Yale University, September 1977, mimeo.
126. William Sewell and Robert Hauser. *Education, Occupation, and Earnings*. New York: Academic Press, 1975.
127. Richard Silkman. "The Diffusion of Educational Innovation." Ph.D. dissertation, Yale University, in progress.
128. Marshall Smith. "Equality of Educational Opportunity: The Basic Findings Reconsidered." In *On Equality of Educational Opportunity*, eds. F. Mosteller and D. P. Moynihan. New York: Random House, 1972.
129. Lewis Solmon. "The Definition and Impact of College Quality." In *Explorations in Economic Research*. New York: National Bureau of Economic Research, Fall 1975.
130. Michael Spence. "Job Market Signalling." *Quarterly Journal of Economics* (August 1973).
131. Joseph E. Stiglitz. "The Theory of Screening, Education, and the Distribution of Income." *American Economic Review* 65 (June 1975).
132. Susan B. Stodolsky and Gerald Lesser. "Learnings Patterns of the Disadvantaged." *Harvard Educational Review* (Fall 1967).
133. Anita Summers and Barbara Wolfe. "Equality of Educational Opportunity Quantified: A Production Function Approach." *Philadelphia Federal Reserve Papers*, 1974.
134. ———. "Do Schools Make a Difference?" *American Economic Review* 67 (September 1977).
135. P. A. V. B. Swamy. "Criteria, Constraints, and Multicollinearity in Random Coefficient Models." *Annals of Economic and Social Measurement* (October 1973).
136. Paul Taubman and Terence Wales. *Higher Education and Earnings*. New York: McGraw-Hill, 1974.
137. Michael K. Taussig. "An Economic Analysis of Vocational Education in the New York City High Schools." *Journal of Human Resources* 3 Supp. (1968): 59-87.
138. Lester Thurow. *Generating Inequality*. New York: Basic Books, 1975.
139. Lester Thurow and R. E. B. Lucas. *American Distribution of Income: A Structural Problem*. Joint Economic Committee, 92nd Congress, March 1972.
140. U.S. Commission on Civil Rights. *Racial Isolation in the Public Schools*, Vols. I and II. Washington: U.S. Government Printing Office, 1967.
141. U.S. Senate. *Hearings Before the Select Committee on Equal Educational Opportunity*, Pts. 1-21. 92nd Congress, 1971.
142. Paul Wachtel. "The Effect on Earnings of School and College Investment Expenditures." *Review of Economics and Statistics* (August 1976).
143. Burton A. Weisbrod and Peter Karpoff. "Monetary Returns to College Education, Student Ability, and College Quality." *Review of Economics and Statistics* (1968).
144. Randall D. Weiss. "The Effect of Education on the Earnings of Blacks and Whites." *Review of Economics and Statistics* (May 1970).
145. Finis Welch. "Measurement of the Quality of Schooling." *American Economic Review* 56 (May 1966).

146. ———. "Education in Production." *Journal of Political Economy* 78 (January/February 1970).
147. ———. "Black-White Differences in the Returns to Schooling." *American Economic Review* 63 (December 1973).
148. David Wiley and Annegret Harnischfeger. "Explosion of a Myth: Quantity of Schooling and Exposure to Instruction, Major Educational Vehicles." *Education Researcher* (March 1974).
149. Donald R. Winkler. "Educational Achievement and School Peer Composition." *Journal of Human Resources* 10 (Spring 1975): 189-204.
150. ———. *The Production of Human Capital*. New York: Arno Press, 1977.
151. Willard Wirtz and others. *On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline*. New York: College Entrance Examination Board, 1977.
152. Kenneth I. Wolpin. "Education and Screening." *American Economic Review* 67 (December 1977): 949-58.

Rendigs Fels*

Multiple Choice Questions in Elementary Economics

The mark of maturity in a science is rigorous hypothesis-testing. Economic education is a new field in the sense that only a little hypothesis-testing has been done so far. This is why it is an exciting field—it is full of opportunities.

Hypothesis-testing requires measurement, quantification, and data. One reason there has been little hypothesis-testing in economic education has been the lack, until recently, of objective measuring instruments.

Important as rigorous hypothesis-testing is for the advancement of economic education, hardheaded evaluation by each teacher of his own accomplishments could also contribute a great deal. College teachers of economics, like the government, have commonly measured their accomplishments by their input instead of their output. In GNP data, the Department of Commerce estimates the output of the government sector by the value of the inputs rather than by the value of the output because there has been no alternative. College economics teachers have judged themselves and their fellows by the amount of sophisticated economics they have put into their courses, not by how much the students have gotten

*I am indebted to Paul L. Dressel, George P. Hollenbeck, and Allen C. Kelley for helpful comments on the first draft of this paper. None of them has seen the final draft, and no one is to blame for the shortcomings of the paper except myself.

out of them, because there has been no alternative. College teachers who deplore casual empiricism in research have relied on nothing else in evaluating teaching.

In the last few years, this situation has changed rapidly. There are now available at least three good multiple choice tests useful for compiling data, testing hypotheses, and evaluating teaching.¹ Since most economists know little about the principles of good test construction, the first section of this paper will summarize the main points. To illustrate them, the second section discusses some multiple choice questions written by the participants in the Stanford Seminar on New Developments in the Teaching of Economics. The third section will present some data derived from the Test of Understanding in College Economics.

I

The most important principles governing construction of objective tests—particularly tests to be published—are these:

1. The first step is to draw up specifications. Normally, specifications for a test consist of a two-way classification giving the percentages of questions to be asked. One classification is by subject matter. For example, in the College Board's test in introductory economics, 40 percent of the questions are on macroeconomics, 40 percent on microeconomics, 19 percent on international trade, and 10 percent on comparative economic systems.² The other classification specifies the type of question or type of material. For example, the specifications for the College Board's test state that 15 percent to 20 percent of the questions test "Ability to apply simple models and use analytic tools."³

¹The Test of Economic Understanding (for high school students), published by Science Research Associates for the Joint Council on Economic Understanding, 1964; the Subject Examination in Introductory Economics of the College Entrance Examination Board's College-Level Examination Program (for the end of a full-year college course); and the Test of Understanding in College Economics (described below).

²As two of these categories are very broad, they presumably were broken down into such subcategories as money and banking, national income analysis, and so on, in the actual work of building the test.

³Study of the questions actually used in the test has shown that "apply" refers to questions with hypothetical rather than real data, to artificial rather than to genuine situations. The College Board's test contains almost no questions of the kind discussed in Section II and the Appendix to this paper.

2. Multiple choice questions are better than true-false questions. Except for unduly easy questions, it is much more difficult to write true-false questions for which a group of experts will agree on what the right answer really is. Moreover, true-false questions encourage guessing. Because there are only two choices, the options do not present a continuous line of thought and do not force the student into clear-cut discriminations. True-false questions therefore reveal less about what goes on in the minds of students. The incorrect answers to multiple choice questions can be written to reveal the nature of the incorrect thinking. (Multiple choice questions normally have either four or five options.) My experience indicates that finding a fifth option that is both plausible and wrong is more trouble than it is worth.

3. Questions must be tried out on a large number of students to weed out those that do not work. "Large number" means enough to give reliable results—preferably several hundred. Data is needed to show that (a) the proportion of students getting the right answer is in the appropriate range (questions that all get right or all get wrong are not useful for discriminating among students); (b) there is positive correlation—preferably a coefficient of 0.30 or better—between right answers to a given question and an index of the quality of the students; and (c) all of the wrong answers are selected by some students ("distractors," as wrong responses are called, that are never chosen make the test inefficient in the sense that testing time is used in a way that contributes no data useful for discriminating among students).

4. All questions must be carefully edited by both an economist and a psychometrician. Editing is difficult, demanding, and time-consuming. The economist must make sure that the right answer really is right and that every one of the wrong answers is indisputably wrong.⁴ The psychometrician must make sure that the question conforms to high standards of test construction from a technical point of view. (For example, testing experts look for "giveaways"—language that enables a test-wise student to detect the right choice without knowledge of the subject.)

⁴If the student is to be presented with two correct options, a key must be used to prevent ambiguity. (Psychometricians frown on the device of making the last option "all of the above" and similar expedients since the preceding options are not unambiguously wrong.) The two correct choices can be labeled I and II in the key and the options made to read: 1. I only. 2. II only. 3. Both I and II. 4. Neither I nor II.

5. If the user wants to know more than just the relative ranking of a group of students, the test in final form needs to be administered to hundreds or thousands of students representing a good cross-section of the population we are interested in to provide norming data.

6. A good multiple choice test in economics can measure a substantial part of student attainment but not all of it. In this respect, economics is somewhere between mathematics and history. The scores on good multiple choice tests in mathematics correlate so highly with grades on other forms of examination that for all practical purposes the correlation can be regarded as 1.0.⁵ In economics, the correlation coefficients may be expected to range between 0.60 and 0.75. In a subject like history, the correlation is lower. To do justice to individual students of economics, course grades should not be based solely on objective tests. Essay tests are needed also. But multiple choice tests can be good enough to provide useful data for hypothesis-testing and evaluation of teaching practices, provided the data is interpreted with the same awareness of its limitations as is required in using the very imperfect data on gross national product and price levels that economists use in their daily work.

7. Writing good multiple choice questions is difficult. It requires an aptitude that can be developed with practice but is not uniformly distributed among economists. The advantages of specialization and division of labor apply to writing multiple choice questions—more so than to essay tests. Essay questions are easy to write but hard to grade. Objective questions are easy to grade but hard to write.

8. Routine questions calling for "book" answers and artificial questions using hypothetical data are easier to write than questions requiring application of economic principles to new, realistic situations, including applications to policy questions.

To the above list of well-established generalizations on construction of multiple choice tests, let me add two helpful hints. One is that multiple choice questions can be extremely useful as a teaching device. An example is their use at the University of Wisconsin, as reported by Allen Kelley, for weekly diagnosis of student progress

⁵ Of course, the computed correlations are lower, partly because of sampling errors.

and tailoring assignments of individual students to test results.* In addition, questions on which the class as a whole had mixed or poor results can be used in lectures and class discussion to clarify points that have not gotten across.

The other helpful hint is that a multiple choice test in which the student is required to give a brief explanation of his answer can be a useful compromise that avoids the most serious drawbacks of essay tests and ordinary multiple choice tests. Most teachers are not in a position to construct good multiple choice tests themselves, since they are unable to try the questions out on hundreds of students before using them or to hire a psychometrician to help edit them. On the other hand, essay questions are subject to wide variability in grading. Essay questions should be graded independently by at least three economists, and the grades averaged, a procedure that is usually impractical. But if the student is asked a multiple choice question and must explain his answer briefly, any deficiencies in the question itself can be detected and allowed for, and the teacher gets quasi-objective data on how much or how little the students have learned.

II

In June 1968, the professors participating in the Seminar in New Developments in the Teaching of Economics took the Test of Understanding in College Economics. There followed a lengthy discussion of the results, which are reported in the next section of this paper. Suffice it to say for the moment that the extraordinary skill economists generally have for detecting flaws, real or imagined, in other people's multiple choice questions was exhibited to the full on this occasion. The participants were then challenged to try to do better themselves. They were given four genuine quotations and asked to write one or more multiple choice questions on them or on a genuine quotation or realistic situation of their own choosing. A modest prize of \$25 was offered for the best question.

Twenty-three of the forty participants in the seminar submitted a total of twenty-six questions. Eighteen of the questions were

* Allen C. Kelley, "An Experiment with TIPS: A Computer-aided Instructional System for Undergraduate Education," *American Economic Review*, May 1968, pp. 446-57.

based on the quotations I supplied, the other eight being entirely original.⁷ Three of the participants submitted a total of four questions that seemed to me reasonably good. Some others had possibilities but needed considerable reworking. The best questions were based on two of the quotations I furnished, with two of the wholly original questions having possibilities.

In interpreting these results, it must be borne in mind that the time available for writing questions was extremely short. The contest was announced late one afternoon. The deadline for submitting entries was the next morning. On the other hand, the professors were to some extent a select group with a special interest in teaching. Moreover, the questions were judged by the loose standards appropriate for items that had not been edited or tried out on students.

Subject to such qualifications, the results accorded with expectations. Questions requiring application of economic principles to genuine quotations or situations are hard to write. I doubt if as many as 10 percent of all economists have any real aptitude for such work.⁸

The prize-winning question was submitted by C. G. Williams of Boston College. It was based on the following quotation (from an economist whose identity I prefer not to publish for reasons that will be obvious):

Some decades ago clean laundry was produced in the household by means of hand laundering. Later, toward the end of the 19th century as mechanical washers were developed, this activity moved into commercial laundries, in the business sector. Presently the movement is back the other way, from the business sector to the household, where homelaundrying devices are rapidly becoming commonplace. The first move was a simple case of mechanization and realization of economies of large scale production. The act of laundering became economically more efficient because machines took over from human drudgery. In the second move, however, economic efficiency was lost because home washing and drying machines operate on a small scale and remain idle during much of their useful life.

⁷Original in the sense that the participant found the quotation or situation himself. The terms of the contest prohibited artificial situations or "quotations" invented for the occasion.

⁸When I told a member of the Committee for a College-Level Test of Economic Understanding that the results of the contest accorded with my expectations, he said I was cynical. I disagree. To say that few economists can write multiple choice questions of a realistic kind is no more cynical than to say that most economists are not econometricians.

Williams' question is reproduced here without any editing.

Is this quotation essentially correct or incorrect and why?

- (a) Correct because it means that housewives waste a lot of effort doing laundry when such work can be more easily done by commercial laundries.
- (b) Correct because large-scale operations, especially in a business like laundering, are always cheaper, *per unit of output*, than small-scale operations.
- (c) Incorrect because, as with home appliances, commercial laundry machinery is not worked continuously and similarly is subject to depreciation and obsolescence.
- (d) Incorrect because the process has taken place in accordance with the preferences of the household regarding the allocation of household time, effort, and income.

The question has the merit of testing the ability of students to detect an error in the use of an important concept: economic efficiency. Its shortcomings are remediable. The quotation would have to be shortened, unless other questions were to be based on it, because it is much longer than necessary. The question preceding the options should refer to the correctness of the last sentence of the quotation rather than to the whole of it. The word "allocation" in option (d) needs to be replaced by "use" since such a favorite word of economists may tip off the right answer to test-wise students. Since response (c) could be defended as possibly correct, though clearly inferior to (d), it might be desirable to specify that the student is to choose the *best* response. Some revision of language and punctuation would be routinely made in the process of editing.

At the time the questions were judged, there was no way to know whether the one by Williams would work with students. Subsequently, I tried it out on twenty-six schoolteachers who had just been through the equivalent of a course in elementary economics at Georgia State University.⁹ None chose (a), seven chose (b), one chose (c), seventeen (two-thirds of the group) chose the right answer, (d), and one did not answer the question. The results suggest

⁹ On the Test of Understanding in College Economics, Part I, their mean score did not differ significantly from the national average of college students. The laundry question was administered in an undesirable way. It was shown on a screen rather than duplicated and distributed. Some participants were too far from the screen to be able to read it clearly. The results may have been adversely affected in consequence.

that the question is in the right range of difficulty. Option (b) is an effective distractor (and shows how badly oversold Americans are on the economies of large-scale production). There is reason to fear that options (a) and (c) are ineffective, but a larger sample would be needed to arrive at a judgment.¹⁰

Later, I gave the question to 150 students at Vanderbilt University at the end of the first week of economics instruction. The students were required to give a brief explanation of their answers. Although a very large majority chose the correct option, all the distractors proved effective. The explanations were generally disappointing. Whether the students chose the right response for the wrong reason, or whether they lacked the skill needed to communicate successfully, or whether they were given too little time is unknown. The fact that the question was given so early in the course undoubtedly accounts in part for the poor explanations.¹¹

Some of the other questions submitted in the competition are discussed in the Appendix to this paper.

III

Table 1 compares the scores on all four forms of the Test of Understanding in College Economics of the forty participants in the Stanford seminar (called "professors" in the table) with the norming data based on a nationwide sample of students. Since the TUCE has been discussed at length elsewhere, only a brief description will be given here.¹² Construction of the test was proposed by the Committee on Economic Education of the American Economic Association and carried out by the Joint Council on Economic Education through a committee of economists and psychometricians and The Psychological Corporation.¹³ There are two forms (called A

¹⁰ No data were collected on the relation between right answers to this question and the ability of the participants. Verification that the correlation coefficient is significantly higher than zero would be needed before the question could be accepted for use in a multiple choice test.

¹¹ The laundry question was preceded on the quiz by the question "What is meant by 'economic efficiency'?" About one-quarter of the answers to that question were given a grade of F.

¹² Besides the Manual for the TUCE, see Rendigs Fels, "A New 'Test of Understanding in College Economics,'" *American Economic Review*, May 1967, pp. 660-66; and Arthur L. Welsh and Rendigs Fels, "Performance on the New Test of Understanding in College Economics," *ibid.*, May 1969.

¹³ The members of the committee were G. L. Bach, William G. Bowen, Paul L. Dressel (Executive Director), Rendigs Fels (Chairman), R. A. Gor-

BEST COPY AVAILABLE

Multiple Choice Questions in Elementary Economics

TABLE 1. Comparison of Scores on Test of Understanding in College Economics (percent correct)

Question Number	Part I (First Semester)						Part II (Second Semester)					
	Form A			Form B			Form A			Form B		
	Pro-fessors	Stu-dents	Differ-ence	Pro-fessors	Stu-dents	Differ-ence	Pro-fessors	Stu-dents	Differ-ence	Pro-fessors	Stu-dents	Differ-ence
1	98	87	11	100	94	6	97	78	19	100	85	15
2	98	85	13	98	79	19	92	84	8	100	88	12
3	98	80	18	100	78	22	82	61	21	100	80	20
4	98	66	32	98	69	29	97	95	2	92	43	49
5	100	72	28	100	81	19	90	83	7	92	77	15
6	95	85	10	98	76	22	92	72	20	100	84	16
7	98	79	19	98	80	18	100	85	15	87	56	31
8	93	74	19	100	65	35	83	76	7	97	81	16
9	98	63	35	85	63	22	87	65	22	83	68	15
10	95	65	30	98	68	30	90	74	16	87	66	21
11	55	58	-3	95	49	46	80	73	7	100	66	34
12	63	45	18	100	69	31	100	73	27	100	76	24
13	50	47	3	88	59	29	92	74	18	97	51	46
14	65	48	17	90	56	34	92	67	25	90	59	31
15	98	60	38	98	53	45	90	58	32	87	56	31
16	98	52	46	98	59	39	92	53	39	69	57	12
17	90	55	35	85	70	15	92	54	38	95	58	37
18	98	58	40	88	53	35	87	48	39	97	54	43
19	93	52	41	95	48	47	74	61	13	77	56	21
20	98	63	35	95	54	41	92	57	35	85	63	22
21	93	50	43	70	47	23	92	49	43	100	60	40
22	90	51	39	78	40	38	69	52	17	77	47	30
23	58	37	21	88	59	29	23	13	10	87	34	53
24	80	47	33	88	50	38	92	39	53	87	41	46
25	68	39	29	90	28	62	82	60	22	87	49	38
26	88	38	50	88	49	39	80	44	36	100	39	61
27	93	45	48	83	53	30	54	36	18	82	34	48
28	88	45	43	98	67	31	90	45	45	87	51	36
29	88	47	41	70	54	16	80	46	34	77	52	25
30	98	46	52	95	39	56	64	34	30	80	30	50
31	93	48	45	83	31	52	36*	19*	17*	90	41	49
32	95	42	53	80	31	49	90	32	58	72	33	39
33	93	32	61	88	25	63	56	26	30	90	45	45
Average	88	56	31	91	58	33	82*	57*	25*	90	55	35

* Scores on question 31 of Part II, Form A low because of an error on the preliminary forms. For students, the estimated norm on the corrected question is 41 percent.

Sources: Students—Manual for Test of Understanding in College Economics (New York: The Psychological Corporation, 1968), p. 15; professors—Stanford Seminar in New Developments in the Teaching of Economics (see text). The averages (means) shown here for students were computed by averaging the (rounded) data in the TUCE Manual (p. 18) and differ slightly from the means given by *ibid.*, p. 18, which were computed from unrounded data.

and B) for the content of each semester of the typical elementary college course in economics. (The forms for the first semester, called Part I, are mainly on macroeconomics; those for the second semester, called Part II, cover microeconomics, comparative economic systems, and international economics.) The specifications determined by the committee conformed, with respect to subject mat-

don (1965-67), B. F. Haley (1967-68), Paul A. Samuelson, and George J. Stigler. John M. Stalnaker served as consultant.

ter content, as closely as possible to current teaching practice; but, with respect to the kind of question asked, the committee decided on a departure: only one-third of the questions were to test "recognition and understanding," with one-third requiring simple applications of economic knowledge, the other third complex applications. In consequence, the TUCE closely resembles the College Board's test only in subject-matter coverage. The TUCE has fewer textbook-type questions, more requiring applications to realistic situations or genuine quotations.

Table 1 verifies that the TUCE is a valid instrument in the sense that those who know more economics get higher scores. In the national sample, students who had received college instruction in economics averaged about 19 right out of 33 questions on all four forms, whereas the participants in the Stanford seminar scored 29 on three of the forms and only slightly less on the fourth. It is evidently a hard test, suggesting that it might be useful at a more advanced level than the elementary course.¹⁴

More interesting than the over-all results are the scores for individual questions. In 124 of the 132 questions, the percentage of professors getting the right answer was at least 10 points higher than for the students. In five of the other eight cases, the professors scored less than 10 percentage points higher largely because the question was easy. For example, 95 percent of the students answered question 4 of Part II, Form A, correctly. The greatest possible margin for the professors over the students was 5 percentage points; the actual margin was 2 points. The remaining three questions (nos. 11 and 13 on Form A of Part I and no. 11 on Form A of Part II) will be discussed individually in a moment. On 111 questions, 80 percent or more of the professors chose the right response. The 13 questions on which they scored less than 70 percent (which include two questions for which a low or negative margin of improvement over the students cannot be explained on grounds of easiness) merit individual attention.

For one of the 13 questions (namely, II-A-31), a key sentence had been omitted from the stem in the preliminary forms used for

¹⁴ A mean score for students in the vicinity of 57 percent is too low for purposes of giving course grades under the traditional scale of 70 percent for a C, 80 percent for a B, and so on, but is just about right for a test designed to be used for evaluation and research at a wide variety of colleges, including the best.

both professors and students. The error has been corrected in the final form as published by The Psychological Corporation.¹⁵

II-A-23 is the only other question on which the professors scored less than 50 percent. This question is on marginal productivity analysis. Hypothetical data given in the stem imply that the firm could increase profits by increasing the use of both factors of production. Apparently professors and students alike are so used to thinking in terms of factor substitution in the context of marginal productivity that they lost sight of profit maximization.

In Part I, Form A, question 11 is based on a genuine quotation from a popular magazine. The analysis implicit in the quotation is faulty, and the question was designed to test the ability of the student to see through it. Although the question does not discriminate effectively between professors and students, the results are interesting. Students tested before studying economics did virtually as well (57 percent) as at the end of the first semester (58 percent). Moreover, this is the only question on which the professors did worse (55 percent) than the students. There are two possible explanations. The question may be faulty, or economics training may contribute little to reading popular magazines critically, or—more likely—both. One of the members of the Stanford Seminar argued that one of the wrong responses could be defended as correct. Since only 10 percent of the professors (and a similar proportion of the students both at the beginning and end of the first semester of economics) chose that response, the shortcomings of the question do not seem to be the major part of the explanation. Although evidence from one question means little, the results suggest what I believe to be the case: that teachers of elementary economics rarely make any effort to train their students to use their knowledge of economics in reading newspapers and magazines. Consequently, neither students nor teachers acquire skill along these lines, and the popular press can get away with publishing bad economics.

¹⁵ I am indebted to the members of the Stanford Seminar for pointing out the error in time to make the correction. It was possible to estimate the national norm for students from tryout data. The estimate is 41 percent, rather than the 19 percent shown in Table 1. The data in Table 1 for this question are based on what would have been the right response if there had been no error in the stem. On the basis of what I consider to be the correct answer to the question as actually asked, 51 percent of the professors and 46 percent of the students got it right. Such figures mean little except to indicate that even with the right key the question as asked was faulty.

Questions 12, 13, and 14 of Part I, Form A, are all based on a long quotation from a newspaper story. In each case, the professor's score was low and, in one case (no. 13), it was only slightly higher than the students' score. In each case, the students did better after a course in economics than before, and the correlation coefficients between this item and scores on the test as a whole were acceptably high. Thus the questions contribute toward discriminating between those who know less and those who know more—but not as much as one could wish, illustrating the difficulty of writing this kind of question.

Question 23 of Part I, Form A, is a straightforward item asking which of four limitations on Federal Reserve policy is most serious for combatting recession. Students do markedly better on this question after taking economics than before. The professors scored markedly higher than the students, but not very high—58 percent. According to the key, the right answer is lack of control over the velocity of money. The most popular incorrect response among both professors and students was inability to control the level of demand deposits. The inference seems to be that although some professors teach and use the velocity concept, quite a few do not.¹⁶

Question 25 of Part I, Form A, gives a short genuine quotation from a news source attributing a large rise in the money supply "primarily" to federal deficits. It is followed by some facts showing that (1) the federal deficit in the period in question was only about a third as large as the excess of business spending over sales and (2) Federal Reserve credit rose sharply. The question asks if the facts "support" what the quotation says about the rise in the money supply and why. The exact word used is crucial to the validity of the question. The facts do not "support" the conclusion, since they suggest that the major source of the increased quantity of money was bank credit to businesses. Of course, the facts do not necessarily contradict the quotation, since the federal deficit might have been the primary stimulus to increased demand for and supply of credit. That only two-thirds of the professors got the question

¹⁶ Despite some argument at the Stanford seminar, I do not believe there can be any doubt about which of the two limitations discussed in the text above is the more serious. Among students, inability to control the outflow of gold was almost as popular a response as inability to control the level of demand deposits and was chosen more frequently after a semester of economics than before—an illustration of how a wrong answer can reveal increased understanding.

right indicates that the question is hard, requiring careful reading and analysis. It is further evidence that teachers make little effort to train students in evaluating popular statements about economics. But this time, as one would expect, the professors did better than the students. The question effectively discriminates degrees of knowledge of economics—25 percent right for students before taking college economics, 39 percent after one semester, 68 percent for professors, and a correlation coefficient for students between this question and total score on the test well above the acceptable minimum.

The scores on no. 11 of Part II, Form A, make it appear easier than the average TUCE question. Nearly three-quarters of the students and four-fifths of the professors got it right. Yet the professors scored only 7 percentage points higher than the students out of a possible 27 points. The question asks about the conditions under which a rise in wages will cause a substantial fall in employment. The three distractors all name conditions that tend to limit the effect on employment. The fourth is of the "none-of-the-above" type. Since elimination of each of the three distractors requires a fair amount of analysis, the professors' score is not surprising in view of the time limit. The puzzle is how the students managed to do so well. One suspects that a good many baffled undergraduates resorted to "none-of-the-above" at a guess.

In Part II, Form A, question 22 reproduces from a newspaper column an estimate of the cost of operating a used car for a year, broken down into major categories. There are two errors in the method of calculating costs, doublecounting (both depreciation and the purchase price of the used car are included) and omission of an opportunity cost (interest on the owner's capital). All the professors spotted at least one of the errors, but only 70 percent saw them both, compared to 52 percent for the students.

Question 27 of Part II, Form A, requires a difficult application of economic theory to a realistic situation. Hard though the question is, it discriminates effectively, the professors scoring 53 percent compared to 36 percent for the students.¹⁷ In this kind of question, it is easy for well-qualified people to get the direction of change wrong, especially under pressure of time. Question 30 of Part II,

¹⁷ Unfortunately, for Part II we do not yet have data on how well students perform at the beginning of the semester, but the correlation between II-A-27 and the total score on the same form is acceptably high.

Form A, is (like II-A-23) a numerical example, this one on marginal revenue. The professors scored 64 percent compared to 34 percent for students. Question 33 on the same form (professors: 56 percent; students: 26 percent) is a difficult abstract question on the effect of a deficit on current account in the balance of payments on interest rates and prices. Most likely, the time limit affected the scores of professors more than the scores of students on all such questions. Professors have a good enough grasp of economic analysis to be able to make effective use of more time to think the problem through and correct initial errors.

Question 16 of Part II, Form B, is one of a series based on a long (genuine) quotation from a newspaper. Only 66 percent of the professors got it right compared to 57 percent for the students. For the students, the correlation between this item and total scores on the form as a whole is marginal but acceptable. In this case, the professors' superior knowledge together with the time limit interfered with their scoring better relative to students. The question has to be carefully stated to cover all the contingencies a trained economist is apt to think of, but there was not time for the professors to consider them all and see that the loopholes had been plugged. As a result, 18 percent of the professors (24 percent of the students) chose the response stating that the question could not be answered without more information.

What conclusions can be drawn from this discussion of individual questions? My first reaction was to fear that the low scores by professors on certain items indicated bad questions. But this turns out not to be the case. Nearly all the questions contribute toward discriminating degrees of knowledge. None detracts from it. Some—notably I-A-11 and II-A-23—contribute very little and for the sake of efficiency are prime candidates for replacement if, as I hope, a revised edition of the TUCE is brought out some day. But even I-A-11 and II-A-23 tell us something of interest.

Rather than reflecting on the quality of the test, the comparison between professors and students in part confirms what we have always known: that economics is difficult. The fact that questions written for students in the elementary course sometimes are hard even for their professors is evidence tending to confirm the hypothesis that the elementary course typically is overloaded. We try to do too much; we expect too much of our students.

On the question of how we train college students to be intelligent

citizens, voters, and molders of opinion on economic policy, the evidence from the TUCE is at best no more than suggestive.¹⁸ But, as indicated above, a number of questions on the TUCE hint that we neither try nor succeed in training students to use economic principles to evaluate critically what they read in the popular press.

Appendix

Two of the better questions written by members of the Stanford Seminar in New Developments in the Teaching of Economics (see Section II above) were based on the following quotation:

Research activity engaged an 18 percent larger number of economists in 1966 than in 1964, as compared with an increase of only about 7 percent in the number of economists engaged in all activities. In spite of the influx of new personnel, the median salary of those engaged in research rose slightly more than average for economists as a whole. (N. Arnold Tolles, "The Economic Status of American Economists, 1966: A Preliminary Report," *American Economic Review*, December 1967, p. 1316.)

W. C. Bonifield of Wabash College based the following question on it:

"The seeming paradox of a relative increase in research economists concurrent with a rising salary paid to research economists relative to all economists is best explained by the fact that

1. research economists produce an easily measured output while other economists do not.
2. much income paid to non-research economists is in the form of fringe benefits.
- *3. the demand for research economists is also rising relative to the demand for all economists.
4. research economists constitute only a small proportion of all economists."

¹⁸ The multiple choice format, in my opinion, cannot adequately measure the ability of a student to combine economic analysis, value judgments, and careful consideration of tradeoffs in making up his mind on a policy question. Even for the more limited task of measuring ability to detect good economics from bad in newspapers and magazines, low scores on a group of multiple choice questions may mean merely that the examiner chose to make the questions hard. It also could reflect a high degree of sophistication in the newspapers and magazines themselves—they may make only sophisticated errors. But that is too good to be true.

C. G. Williams of Boston College based the following question on the same quotation:

"This is evidence that

- a) in the market for economists an increase in the supply of research economists (an outward shift of their supply curve) tends to reduce the relative median salary of economists in non-research activities.
- *b) over the period as a whole there was an excess demand for research economists that was met partly by an increase in their number and partly by an increase in their salaries relative to economists engaged in all activities.
- c) the demand for research economists tended to be positively sloping because despite the fact that their salaries rose relative to economists engaged in all activities, relatively more economists were able to find employment doing research work.
- d) among economists now entering the profession, research work carries prestige and status which means that the more able, the better trained, and hence the higher paid economists, actually prefer research positions."

The questions asked and the correct answers are essentially the same in the two items, but the distractors are different. Bonifield's version has the merit of conciseness, but Williams' could be cut during editing. Although only a tryout on students could determine whether the questions would work in practice, they look promising. They test the ability of a student to use supply-and-demand analysis to solve a problem.¹⁰

One of the quotations given to the students as a basis for writing questions was from a syndicated column, "Ask Andy" (the quotation appeared, among other places, in the *Nashville Tennessean*, April 16, 1968, p. 16): "Every dollar bill must have value—and its worth is backed by the government, so the number of bills printed must be limited by the wealth of the country in gold and money, in lots of other assets and resources." It is interesting that nobody succeeded in writing a good question on this bit of nonsense.

The best of the questions based on a quotation or situation other than those furnished to the participants was submitted by William P. Kinney of Foothill College:

¹⁰ Bonifield's question was the runnerup in the contest described in Section II of this paper. My decision in favor of Williams' question on commercial vs. household laundering was based primarily on the belief that it is more important to test whether students understand what economic efficiency is all about than to test their skill in supply-and-demand analysis, important as the latter is. Economists who are more interested in equipping students with the tools of their trade may disagree with me. The fact that Williams submitted a second question with merit, though irrelevant, strictly speaking, to the awarding of the prize, reinforces my conviction that justice was done. Whether or not the decision of the judge is final.

Multiple Choice Questions in Elementary Economics

"But the Vietnam war has had its impact on health programs for our citizens. Medicaid expenditures—which provide assistance to the medically needy of whatever age—were restricted last year because of the budgetary squeeze." (From *"First Things First"* by Senator Eugene J. McCarthy.)

The above quotation best illustrates:

- A. The inefficacy of fiscal policy in the Vietnam war years.
- B. The concept of opportunity cost.
- C. The clear need for a more expansionary monetary and fiscal policy.
- D. The need for greater "fiscal responsibility" on the part of the government.

A psychometrician would probably criticize the question for lack of parallelism between the right answer and the distractors. The right answer is a concept, whereas the distractors are characterizations of policies. From the point of view of substance, the question is in part a test of vocabulary. It is more important for a student to be able to use a concept like opportunity cost than to know the name economists have given it. But the question tests ability to recognize the idea in an unfamiliar context, and one can argue that the concept is so important that students should know its name.

The Test of Understanding In College Economics and Its Construct Validity

Darrell R. Lewis and Tor Dahl

What does the new Test of Understanding in College Economics (TUCE) actually measure, and of what significance is this to those interested in evaluating student performance in the introductory college economics course? Although some results obtained during the norming [11,6,17] and early experimental use of TUCE [1,9,7] have been reported, a number of construct validity and research design questions remain. The purpose of this paper is to present additional data on the TUCE, primarily with regard to its validity as an experimental testing instrument and as to its construct design.

In brief, the results from this study indicate that: (1) TUCE is effective in discriminating between "good" and "poor" students in economics. (2) Although academic ability and critical thinking skills are related to achievement in economics and on TUCE, considerable independence between TUCE and prior ability exists. TUCE incorporates prior ability and critical thinking skills while also effectively discriminating on other knowledge. (3) The subparts of TUCE can be differentiated in that they do seem to measure different things. (4) However, when associating the three subparts of TUCE with critical thinking skills as measured by the Watson-Glaser Critical Thinking Appraisal (CTA), the most significant associations occur with the "simple application" and not the "complex application" types of questions. In fact, the complex application questions in TUCE show minimal association with critical thinking skills. Our results indicate that the researcher using TUCE must be cautious about imputing higher educational value to complex application types of questions.

Darrell R. Lewis and Tor Dahl are Associate Professor of Economic Education and Lecturer of Economics in Public Health, respectively, at the University of Minnesota.

Source of Data and the Test Instruments

Primary data for this paper were obtained from an experimental research study dealing with critical thinking skills in the introductory course undertaken at the University of Minnesota in 1969.¹ For the purpose of this paper, it is only important to know that during the 1969 fall quarter, 784 University of Minnesota students in 23 sections of Economics I (Principles of Economics—Macroeconomics) were subjected to before and after survey questionnaires as well as being pre- and posttested on the TUCE (Part I, Forms A and B) [11] and the Watson-Glaser Critical Thinking Appraisal (CTA) (Forms ZM and YM) [16].

The CTA, which has been nationally normed and validated, has frequently been used as a measure of critical thinking achievement in instructional situations at the secondary or college level and in industrial and executive programs [4,13,18,8]. It has also had extensive usage as a research tool to determine the relationship between critical thinking abilities and other abilities or traits [16,14].

The CTA instrument consists of a series of test exercises requiring the application of some of the more important abilities involved in critical thinking. The exercises include problems, statements, arguments and data interpretations similar to those which an enlightened citizen might encounter in his daily life as he might work, read newspapers or magazine articles, hear speeches, or participate in discussions of various issues. Of three such national instruments, the one selected for this study most closely approximates the criteria for critical thinking skills discussed by Paul Dressell [2], Morse and McCune [10], and the American Council on Education [3]. The test, available in two forms (YM and ZM), consists of five subtests designed to measure different, though interdependent, aspects of critical thinking.² Thus each form contains 100 items that can be completed in about 40 minutes. In some items the student is asked to think critically about problems involving "neutral" topics such as the weather, scientific facts or experiments, and other matters about which people generally do not have strong feelings or prejudices. Other items, approximately parallel in logical structure, pertain to social issues concerning which many people have emotional feelings, biases or prejudices.

The TUCE has been discussed in detail elsewhere [17,15]. Each of its four forms has three subparts of 11 questions involving (a) "recognition and understanding," (b) "simple application," and (c) "complex application." The primary purpose of the test is to provide a research instrument for controlled experiments, and thereby a basis for evaluating different versions of the introductory college economics course [17, p. 224, and 15, p. 62]. Consequently, it is important that we have added data on TUCE as a measurement instrument.

¹The research design and results of that study are reported in another paper and may be obtained by writing to the authors. Darrell R. Lewis and Tor Dahl, *Critical Thinking Skills in the Principles Course: An Experiment*. Unpublished paper, University of Minnesota, 1970.

²The five subtests are as follows:

- Test 1. Inference (20 items). Samples ability to discriminate among degrees of truth of inferences drawn from data.
- Test 2. Recognition of assumptions (16 items). Samples ability to recognize unstated assumptions which are taken for granted in given statements.
- Test 3. Deduction (25 items). Samples ability to reason deductively from statements; to recognize the relation of implication between propositions; to determine whether what may seem to be an implication from premises is indeed such.
- Test 4. Interpretation (24 items). Samples ability to weigh evidence and to distinguish between (a) generalizations from data that are not warranted beyond a reasonable doubt, and (b) generalizations which, although not absolutely certain, do seem to be warranted.
- Test 5. Evaluation of arguments (15 items). Samples ability to distinguish between arguments which are strong and relevant and those which are weak or irrelevant to a particular question.

TUCE as a Discriminating Measure

As Rendigs Fels points out [6, pp. 4-5], we have preliminary evidence that TUCE, as a discriminating measure of performance in the principles of economics, is a valid instrument. "Before studying economics, college students in a national sample got 13 questions right out of 33 (40%). After studying economics, they got 19 right (58%). A sample of professors got 29 right (88%). Thus, those who presumably know more economics score higher than those who know less. Furthermore, there is evidence that every single question works in the sense that good students are more likely to get it right than poor students."³ These results were essentially confirmed in the Minnesota data. Table 1 (column 5) indicates that the Minnesota students increased their scores on TUCE from a mean of 14.16 to 19.85, an increase of over 40 percent. This value-added of 40 percent approximates the improvement factor of 41 percent found in the national averages of the norming data [17, p. 225].

Table 1
TUCE Performance by Selected Groups †

	(1) Honors (N=21)††	(2) Top 27% (N=210)	(3) Middle (N=359)	(4) Bottom 27% (N=215)	(5) Total (2+3+4) (N=784)†††
	$\bar{X}$	$\bar{X}$	$\bar{X}$	$\bar{X}$	$\bar{X}$
Pre-TUCE	19.15	17.59	14.06	11.53	14.16
Post-TUCE	25.93	25.46	19.97	14.96	19.85
Increase	6.79	7.87	5.91	3.43	5.69
Percentage increase	35%	45%	42%	30%	40%

†Differences between pre- and post-TUCE in all five groups were statistically significant at the .001 level.

††The 21 students from the two honors sections are *not* included in either the top 27 percent nor in any of the other Minnesota data reported in this paper.

†††Mean scores for TUCE in Table 1 differ slightly from mean scores in the other tables. In the other tables the number of students included is slightly smaller as a result of incomplete ACT scores on transfer students and because some students failed to complete all of the questionnaires.

Table 1 also indicates that TUCE effectively discriminates performance of varying ability levels. In fact, in both absolute and percentage terms, TUCE discriminates more effectively for the more capable regular students; nor does there appear to be a "ceiling effect" in the use of TUCE with the principles course. Even for students from two honors sections who were not included in the regular total (Table 1, Column 1), TUCE effectively discriminated a value-added of 6.79 points, but since the honors sections produced a pre-TUCE average score of 19.15, this high base resulted in a lower percentage gain than was achieved by the upper 27 percent. Nonetheless, the honors sections still produced the highest average post-TUCE score.

Furthermore, the data indicate that those who scored highest on the pretest were approximately the same ones who scored highest on the posttest. The same held true for the lowest group; low pretest scores were also low scorers on the posttest. Of the lowest 27 percent of students on the pretest (215 students), we found 53 percent, or 115 students, still in the lowest 27 percent of scorers on the posttest. Similarly, of the top 27 percent of students on the pretest (210 students), we found 58 percent, or 121 students, still in the top

³Interestingly, our 13 graduate student instructors for Economics I also took TUCE and our results of an average 29 correct approximate Fels' sample of professors [7].

27 percent on the posttest.⁴

These results elaborate on those from the national norms as reported by Welsh and Fels [17, p. 225]. Although limitations of the national norming data precluded comparisons of good and poor students on a before and after basis, these types of comparison *can* be made on the strength of data we have processed for our 784 students.

TUCE and Content Validity

As pointed out by the American Economic Association's Test Committee responsible for the construction of TUCE, "... to the extent that TUCE measures what the Committee judged important, the Test has *content validity*, and is a valid test" [11, p. 15]. Content validity of TUCE, as defined above, cannot be determined statistically. However, an important measure for each experimenter using TUCE could be the degree to which the students' course grades or other relevant examination scores correlate with post-TUCE performance. As Table 2 indicates, the Minnesota data had a post-TUCE zero-order correlation coefficient of .58 with the student's final grade. It is important to note that the instructors did not teach in order to improve performance on the instrument, nor did they use the post-TUCE scores in computing the final grade. In fact, six of the 13 graduate student instructors expressed opinions concerning the "inappropriate" coverage of TUCE relative to "their" principles course. In the light of these qualifications, it is surprising that the correlation coefficient is as high as it is.

Another method of examining what TUCE measures is to study correlations between performance on TUCE and performance on other tests. Although similar data have been developed by Saunders for College Entrance Examination Board Verbal and Mathematical test scores [11, p. 18], none have been presented for student scores on either the American College Test (ACT) or the CTA.

In our study students took TUCE (Part I, Form A) as a pretest and TUCE (Part I, Form B) as a posttest, and the TUCE scores were correlated with both ACT and CTA test scores. Some results are presented in Table 2. By its very nature, TUCE as a measure of achievement might be expected to show a relatively high relationship to such measures of ability as ACT and CTA. However, in the light of such expected relationships the correlations in Table 2 are 'moderately' low. This substantiates the results found by Saunders in which the highest College Board-TUCE zero-order correlation was .44. It further indicates that although academic ability (as measured by ACT) and critical thinking skills are related to achievement in economics and on TUCE, considerable independence between TUCE and these other abilities also exists. TUCE both incorporates prior ability and critical thinking while also effectively discriminating on other kinds of knowledge.⁵

⁴In separate multiple linear regressions not reported in the text or tables, we found confirmation of our suggestion that the amount of economic knowledge (as measured by TUCE) and that the student brings into the course makes a difference in postcourse performance. When pre-TUCE test scores were included as an independent variable in a multiple regression which included post-TUCE as the dependent variable, the pre-TUCE test scores were highly significant (at greater than a .01 level). In fact, the resulting regression coefficient of .30 indicated that approximately 1 point was contributed to the student's post-TUCE score for every 3 correct scores on pre-TUCE.

⁵In separate multiple linear regressions not reported in the text or tables, we found confirmation for these zero-order correlations and inferences. When pre-CTA test scores were dropped from a multiple regression which included post-TUCE as the dependent variable and pre-TUCE as an independent variable, the pre-TUCE test scores become much more significant with its regression coefficient more than doubling. Similarly, the R^2 dropped by over .10. To some extent, CTA and TUCE are surrogates for one another—i.e., they both measure similar types of "thinking processes" or "skills." However (and as inferred above) they still give individually large residual measurement to other knowledge. Consider, for example, the significant drop in the R^2 as a result of dropping the pre-CTA scores.

Table 2
Zero-Order Correlations of TUCE Part I with Scores on Selected Aptitude Measures
(N = 649)

	<i>r</i>	
	TUCE Form A Pretest ($\bar{X}$ = 14.42, SD = 3.74)	TUCE Form B Posttest ($\bar{X}$ = 19.96, SD = 4.01)
1) Total ACT score† ($\bar{X}$ = 24.10, SD = 3.56)	.41	.44
2) Critical Thinking Appraisal (Form ZM) ($\bar{X}$ = 70.10, SD = 9.24)	.41	.37
3) Critical Thinking Appraisal (Form YM)†† ($\bar{X}$ = 71.32, SD = 7.83)	.39	.38
4) Cumulative grade point average ($\bar{X}$ = 2.55, SD = .52)	.30	.41
5) Grade in the Principles course ($\bar{X}$ = 2.41, SD = 1.07)		.58

†ACT has a zero-order correlation coefficient of .52 for CTA (ZM) and .48 for CTA (YM).

††All post-CTA (YM) scores were converted into pre-CTA (ZM) equivalent raw scores [16, p. 8].

TUCE and Construct Validity

One measure of *construct validity* for TUCE may be logically inferred from the Test Committee's definition of the three subparts as presented in the *Manual* [11]. Moreover, recent research [1] has assumed the existence of identifiable subdivisions on TUCE and has consequently interpreted significant differences in the results on this basis.

Evidence of construct validity can also be obtained from a study of the ways in which the various parts of the test are related to each other and to the test as a whole. The interrelationships of the three subtests and the total test are reported in Table 3 from our data. The moderately low subtest zero-order correlation coefficients, ranging from .33 to .36, support the contention that relatively distinctive abilities are being measured with sufficiently small overlap to warrant their inclusion in one total score.

Table 3
Correlation Matrix on TUCE and Subparts of TUCE, Interaction Terms Included
(N = 784)

	TUCE (Form IB)	X ₁ R	X ₂ S	X ₃ C	X ₄ RS	X ₅ CS	X ₆ CR
Recognition/Understanding (R) X ₁	.75						
Simple Application (S) X ₂	.74	.36					
Complex Application (C) X ₃	.77	.36	.33				
R · S X ₄	.89	.78	.84	.43			
C · S X ₅	.91	.45	.82	.79	.79		
C · R X ₆	.92	.79	.43	.83	.74	.77	
C · S · R X ₇	.96	.71	.74	.72	.91	.93	.90

Table 3 is corroborated by the following regression equation which includes multiplicative interaction terms from our data:

$$\begin{aligned} \text{Post-TUCE} = & 1.73 (\text{constant}) + .82 (R) + .76 (S) + .77 (C) \\ & (9.66) \quad (7.91) \quad (8.88) \\ & + .03 (RS) + .03 (CS) + .03 (CR) - .003 (CSR) \\ & (1.85) \quad (2.41) \quad (2.10) \quad (-1.89) \\ \text{Adj. } R^2 = & .99 \end{aligned}$$

The *t*-values are given between parentheses.

Two of the pair combinations show significant interaction, but both the *t*-values and the size of the corresponding regression coefficients are small compared to the noninteraction terms. Obviously higher level interaction *could* be present, but the high R^2 does not seem to warrant further tests, at least not for predictive purposes. The recognition and understanding term (R) seems to be the most identifiable or "purest" part of post-TUCE; while this distinction is small, it is intuitively acceptable. One would expect application questions to have more transferability or overlap than recognition and understanding questions.

A further measure of construct validity for the three subparts of TUCE can be gained by relating TUCE and its subparts to other types of measuring instruments such as the CTA discussed earlier in this paper. Regressions 1-8 in Table 4 present results from simple linear regressions with TUCE and its subparts on CTA. As both the regression coefficients and intercepts are significantly different for CTA in the regressions with the three TUCE subparts (regressions 3-8), it appears that separation of the three subparts of TUCE is valid in that they can be differentiated when referenced to the critical thinking skills measured by CTA. Although the regression coefficients for CTA with both simple and complex application types of questions (regressions 4 and 5, 7 and 8) have approximately identical values, the intercepts are significantly different.

Table 4
Regression Results with Post-TUCE on CTA
(*N* = 849; *t*-ratios in parentheses)

Dependent Variable	Constant†	Independent Variables†		Adj. R^2 †
		Pre-CTA	Post-CTA	
1. Post-TUCE	3.63	—	.21 (9.67)	.15
2. Post-TUCE	9.48	.15 (9.18)	—	.13
3. Post-Recognition/ Understanding	3.91	.04 (6.66)	—	.06
4. Post-Simple Application	1.73	.06 (9.88)	—	.12
5. Post-Complex Application	2.95	.06 (8.19)	—	.08
6. Post-Recognition/ Understanding	2.26	—	.06 (6.95)	.06
7. Post-Simple Application	—1.11	—	.09 (11.24)	.14
8. Post-Complex Application	— .13	—	.09 (9.96)	.12

†All significant at .05 level.

Notice also the improvement in significance obtained by regressing postvalues with each other, as compared to pairing pre- and postvalues. This is to be expected, as a likely outcome of the intervening educational process. Overall, the application categories of post-TUCE correlate more strongly with CTA than the recognition and understanding terms. This lends weight to the previous discussion of transferability. The recognition and understanding category measures more specific economic knowledge; the simple and complex application categories are perhaps more general in nature.

Adding other significant terms to the regression, we obtain⁶:

$$\begin{aligned} \text{Post-CTA} = & 37.94 - .58 (\text{College Math}) + 3.80 (\text{Reads Newspaper}) \\ & (-2.02) \quad (2.27) \\ & + .57 (\text{ACT Score}) + .23 (\text{Pre-CTA}) + .24 (\text{Pre-TUCE}) \\ & (6.23) \quad (7.56) \quad (3.74) \\ & + .65 (\text{Post-Simple}) + .40 (\text{Post-Complex}) \\ & (4.00) \quad (2.73) \\ \text{Adj. } R^2 = & .39 \end{aligned}$$

The *t*-values are given between parentheses.

Thus, with post-CTA scores as the dependent variable and employing a step-down procedure with a .10 cut-off level, we found that "post-recognition and understanding" types of questions were dropped along with a number of other independent variables. Although "recognition and understanding" did not drop until quite late in the step-down procedure, it confirmed our expectations that these types of questions represent less general conceptual skills as measured by CTA.

It is important to note that both pre-TUCE as a total score and the two post-subparts of "simple application" and "complex application" remained significant during the step-down regression procedure. Furthermore, simple application with a regression coefficient of .65 is a stronger influence on CTA than is the .40 coefficient for complex application—again indicating that simple application types of questions have a higher association with critical thinking skills (as measured by CTA) than do the complex application questions.⁷ The zero-order correlations show a slightly stronger relationship between (R) and (C) than between (S) and (C). Thus, the complex application questions' closer affinity to "pure economics" may remove it from general critical thinking skills more than the simple application questions. This is important to recognize to the extent that some researchers employing TUCE impute a higher general educational value to the complex application types of questions, assuming that they measure a higher (more transferable) reasoning ability. This may be true and they may have higher educational

⁶The significance of, and discussion for, these regression results are given in another paper, "Critical Thinking Skills in the Principles Course: An Experiment." Copies may be obtained by writing to the authors of this paper.

⁷It has been suggested that the lower and less significant regression coefficient for post-complex application questions results from the nature of these questions in that they cannot be answered without some prior knowledge involving "recognition and understanding" items—i.e., they are sequentially dependent upon specific recognition and understanding knowledge. It is further argued that in a multiple linear regression with all three TUCE subparts included, a portion of the "complex" variable is already accounted for and held constant by the "recognition" variable. If these conditions are true, then it is not surprising that the post-complex coefficient is lower and less significant. However, in the multiple regression above, the "recognition" variable has already dropped in the step-down procedure without significantly altering the constant or any of the other regression coefficients. Thus, the total independent (and less significant) condition of post-complex cannot be rejected on this basis.

Table 3
Regression Equations on Subparts of Critical Thinking Skills
(N = 648; t-ratios in parentheses)

Dependent Variable	Independent Variables (Post-CTA Subparts)						Adj. R ² ††
	Constant††	X ₁ Inference	X ₂ Assumptions	X ₃ Deduction	X ₄ Interpretation	X ₅ Evaluation	
Post-TUCE (IB) $\bar{X} = 19.96$	4.96	.147†† (2.38)	.120† (1.47)	.383†† (6.02)	.140†† (2.02)	.077 (.90)	.14
Recognition and Understanding $\bar{X} = 6.90$	3.20	.062†† (2.18)	.014 (.34)	.101†† (3.45)	.031 (.97)	.006 (.14)	.05
Simple Application $\bar{X} = 6.05$	0.18	.044† (1.64)	.067†† (1.92)	.150†† (5.45)	.040† (1.35)	.051† (1.39)	.12
Complex Application $\bar{X} = 7.01$	3.01	.032 (.57)	.021 (.29)	.148†† (2.56)	.025 (.39)	-.021 (-.27)	.02

†Significant at .10 level.

††Significant at .05 level.

value, but any researcher using TUCE will have to rationalize this higher value on the basis of other criteria than those "critical thinking skills" or "reasoning abilities" measured by CTA.

Tables 5 and 6 give additional insight into the types of critical thinking skills which are involved in resolving the problems and questions of the three TUCE subparts.

Table 5
Correlation Matrix for Post-TUCE and Post-CTA Subparts

	TUCE (Form 1B)	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Inference	X ₁	.19							
Assumptions	X ₂	.19	.20						
Deduction	X ₃	.34	.26	.32					
Interpretation	X ₄	.23	.20	.29	.42				
Evaluation	X ₅	.15	.11	.14	.30	.23			
Recognition and Understanding	X ₆	.75	.14	.09	.20	.13	.08		
Simple Application	X ₇	.74	.16	.19	.32	.20	.16	.36	
Complex Application	X ₈	.77	.06	.06	.13	.07	.03	.36	.33
Total CTA (Form YM)		.38						.25	.38
								.38	.34

Note that student performance on "inference," "deduction," and "interpretation" as defined by the CTA are all significantly associated at a 5 percent level with postcourse performance on TUCE (regression 1). For example, 1 point on the "deduction" subpart of CTA contributed .38 point to TUCE. Equally important are the results which show that student performances on TUCE are much less associated with their abilities to discriminate on the basis of "recognizing assumptions" or "evaluating arguments," measured by CTA. However, recognition of assumptions is significantly associated with answering "simple application" types of questions correctly (regression 3). Presumably the nonsignificant association of "assumptions" on total TUCE in regression 1 results from its very low and nonsignificant associations with both "recognition and understanding" and "complex application" types of questions in regressions 2 and 4, thus strengthening its significance with "simple application" from regression 3.

The results of the zero-order correlations from Table 6 confirm these observations. "Deduction" has a correlation coefficient of .34 with TUCE, while ability to recognize assumptions (X₂) correlates most highly with simple application (X₇).

A second observation from Tables 5 and 6 indicates that when we relax the significance level to 10 percent we note that simple application (regression 3) significantly associates with all five CTA subparts. The relatively high correlation coefficients in Table 6 for simple application (X₇) with the five CTA subparts (X₁, X₂, X₃, X₄, X₅) also support the regression results. In every instance, the correlation coefficients for each of the five CTA subparts are highest for simple application and lowest for complex application. This further substantiates the regression results in Table 4 and the multiple regression discussed earlier wherein "simple application" types of questions associated more closely with those abilities measured by CTA than did either of the other two TUCE subparts. Again, it appears that "complex application" types of questions do not

correlate closely with many of the critical thinking skills or attributes popularly associated with such questions [11, p. 7].⁸

A third observation to be noted from Tables 5 and 6 concerns the relative importance and association of "recognition and understanding" types of questions with critical thinking skills as measured by the CTA. Regression 2 significantly associates both inference and deduction abilities to performance on recognition and understanding, while only deduction associates with complex application in regression 4. These results are confirmed further in Table 6, where in every instance recognition and understanding (X_4) has a higher correlation coefficient than complex application with the five subparts of CTA (X_1, X_2, X_3, X_4, X_5).

This significance and association of recognition and understanding types of questions with critical thinking skills confirms the TUCE Test Committee's rationale for this category [11, p. 6]:

Questions in this category need not (and should not) be answerable by rote memory. The better questions of this type involve restatement or recognition of an idea in somewhat different language from that in which it was originally learned. These questions may call for explanation, for summarization, or for simple extension of an idea. Thus, such questions can and should test understanding or comprehension rather than recall.

Clearly, as indicated by the cross validation with CTA, these types of questions can and do test understanding and comprehension rather than simply recall.

However, as the Test Committee appropriately notes, "... the classification of items by objectives has been *a priori*, based largely on the judgment of the cooperating economists and some of their graduate students. Whether the actual mental processes of undergraduate students tackling these questions would correspond to the classifications is not surely known" [11, p. 7]. If (and assuming that) the CTA measures many of the things involved in the kind of "orderly thinking" called for by the National Task Force on Economic Education under the heading, "A Rational Way of Thinking About Economic Problems" [12, pp. 14-17], then we must question whether complex applications do, in fact, measure these objectives [11, p. 7]. Our data indicate that simple application has the most significant association with these objectives. Even recognition and understanding associates more highly with critical thinking skills as measured by CTA than does complex application. Suggestively, complex application types of questions either have to be changed, rationalized on some other basis, or subjected to further testing and scrutiny.⁹

⁸It has been suggested by Henry Villard that the fact that complex application questions do *not* correlate as highly with the CTA as the other TUCE subparts may suggest that they, in fact, have "special educational values and attributes." It is argued that any bright student (as measured by the CTA) can handle simple application and recognition and understanding questions. However, an effective measure of *economic understanding* apart from intelligence (and whatever else the CTA measures) can be indicated by performance on complex application questions. This may be true. However, it again raises the problem of testing the assumption against another validated instrument for measuring whatever constitutes this facet of "economic understanding" or that component which complex application questions measure. Possibly the Graduate Record Examination or the College Level Examination Program in Economics of the College Entrance Examination Board could provide such a vehicle for testing this hypothesis.

⁹In an effort further to identify possible relationships within the 33 TUCE questions and between the TUCE and CTA measures, two factor analyses were performed. The first analysis included only the 33 questions in post-TUCE (Part I, Form B) while the second included the five CTA subparts with the TUCE questions. In both analyses the resulting varimax factor loading matrix (for the eleven and twelve factors, respectively) indicated that very little of either the TUCE content categories or subpart classifications [11, p. 8] could be explained by the factor structures. The only

Concluding Comments

As always in studies like this, one is well advised to resist generalizations based upon a single source of data; clearly the study needs to be replicated. Yet some tentative results do give important insights into the validity of TUCE both as a research and as a measurement instrument.

This study shows that TUCE is not only effective in discriminating between "good" and "poor" students in economics, but that it also measures prior ability and critical thinking skills while effectively discriminating on other knowledge. In addition, the subparts of TUCE, as developed by the Test Committee, can be shown to measure significantly different things. But, when associating the three subparts of TUCE with the critical thinking skills measured by CTA, the most significant associations occur with simple and not complex application types of questions. Our results indicate that the researcher employing TUCE must be cautious about imputing certain educational values and attributes to complex application questions. The recognition and understanding questions, however, were appropriately classified by the Test Committee in that they seem to measure elements other than just memorized items from economic texts. It is suggested that further validation studies should examine the contribution of the complex application questions of TUCE and assess their role in measuring performance in economics.

REFERENCES

1. Richard Attiyeh, G.L. Bach and Keith G. Lumsden, "The Efficiency of Programmed Learning in Teaching Economics: The Results of a Nationwide Experiment," *Amer. Econ. Rev.* 59 (1969), 217-223.
2. Paul Dressel, "Critical Thinking: The Goal of Education," *National Education Association Journal* 41 (1955), 418-420.
3. Paul Dressel and Lewis B. Mayhew, *General Education: Explorations in Evaluation*. Final Report of the Cooperative Study of Evaluation in General Education. Washington, D.C.: American Council on Education, 1954.
4. Robert H. Ennis, "An Appraisal of the Watson-Glaser Critical Thinking Appraisal," *Journal of Educational Research* 52 (1958), 155-158.
5. Rendigs Fels, "A New Test of Understanding in College Economics," *Amer. Econ. Rev.* 57 (1967), 660-666.
6. Rendigs Fels, "Hard Research on a Soft Subject: Hypothesis Testing in Economic Education," *So. Econ. Jour.* 36 (1969), 1-9.
7. Rendigs Fels, "Multiple Choice Questions in Elementary Economics," in Keith G. Lumsden, ed., *Recent Research in Economic Education*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1970.
8. J. Follman et al., "Factor Analysis of Achievement, Scholastic Aptitude, and Critical Thinking Subtests," *Journal of Experimental Education* 38 (1969), 48-53.
9. Campbell R. McConnell and Charles Lamphear, "Teaching Principles of Economics Without Lectures," *Jour. Econ. Ed.* 1 (1969), 20-32.
10. H.T. Morse and G.H. McCune, *Selected Items for the Testing of Study Skills and Critical Thinking*, 3rd ed. Washington, D.C.: National Council for the Social Studies, 1957.
11. The Psychological Corporation, *Manual, Test of Understanding in College Economics*. New York: The Psychological Corporation, 1968.

significant and clearly identifiable factor was the expectedly high clustering of the five CTA subparts in the structure of one factor.

Space limitations preclude reproduction of the matrix in this paper. However, copies may be obtained by writing to the authors.

12. *Economic Education in the Schools: Report of the National Task Force on Economic Education*. New York: Committee for Economic Development, 1961.
13. Velma I. Rust, "Factor Analysis of Three Tests of Critical Thinking," *Journal of Experimental Education* 29 (1960), 177-182.
14. Ibrahim Q. Saadah, "The Teacher and the Development of Critical Thinking," *Journal of Research and Development in Education* 3 (1969), 87-99.
15. Henry H. Villard, "Where We Now Stand," *Jour. Econ. Ed.* 1 (1969), 60-66.
16. G. Watson and E.M. Glaser, *Watson-Glaser Critical Thinking Appraisal Manual*. New York: Harcourt, Brace, and World, 1964.
17. Arthur Welsh and Rendigs Fels, "Performance on the New Test of Understanding in College Economics," *Amer. Econ. Rev.* 59 (1969), 224-229.
18. B.W. Westbrook and J.R. Sellers, "Critical Thinking, Intelligence and Vocabulary," *Educational and Psychological Measurement* 27 (1967), 443-446.

IV
Research Reports

Educational Production Function for an Introductory Economics Course

ELISABETH ALLISON

Harvard has been teaching introductory economics for 162 years: as a branch of moral philosophy for 50 and as a separate discipline for 112. In intellectual terms the core of shared experiences for the 87,000 students who have passed through the course is minimal: Adam Smith's *Wealth of Nations* is the only reading common to both courses. Over time, the course has evolved, following the development of modern economics. Moreover, today's course is not only substantively different from its predecessor; it is demonstrably better. The student who passes Economics 10 today can answer many questions that puzzled students (and indeed their professors) in the 1870s—for example, Adam Smith's paradox of diamonds and water or Malthus's observation of "general gluts."

The contrast between this steady, systematic cumulation of knowledge in the discipline of economics and the noncumulative, essentially personal nature of knowledge about the teaching of economics is striking. In 1973—when the effort described in this paper began—it was difficult to articulate a single proposition about teaching economics effectively that was known to today's instructors but hidden from eight preceding generations of teachers.

The failure to develop a body of substantive propositions about the teaching of economics was mirrored in every facet of the course. When cutbacks in the instructional budget had to be made in 1972, the decision to have larger classes with better-qualified instructors was based on precisely the same

information on which the opposite decision had been made in 1888. No attempt was made to train new instructors because there was no body of knowledge on pedagogy to be conveyed. With no propositions to be tested there was no point to the systematic collection of data on course inputs or outputs.

This contrast was all the more ironic because the process was well suited to empirical investigation. The course had large numbers of students (800–900 per year), rich data from Harvard's elaborate admissions process, well-defined performance standards, and a decentralized course structure that permitted "controlled" experiments. Indeed the situation offered the prospects of data far more reliable than those available for most economic analysis. Moreover there existed—within the discipline of economics—a body of sophisticated research in which the problems of modeling production processes and estimating their parameters had been well worked out and a set of extensions which suggested that this model had general applicability.

Nevertheless, the contrast and the opportunity it posed might well have passed unnoticed for another decade had it not been for two events. The first was a decision to experiment with a new pedagogy, self-paced instruction (SPI).¹ A year's experience suggested that SPI altered the whole learning process, not just a few exam scores. But without a model of the basic process, it was impossible to quantify and evaluate those SPI effects. The second was a warning that

EDITORS' NOTE: At the time of writing, the author was associate professor of economics at Harvard University. Unlike the other papers in this book, this one was not previously published but was circulated as Discussion Paper 545 of the Harvard Institute of Economic Research (April 1977). In preparing the paper for the purposes of this volume, we edited it substantially and thus must share responsibility with the author for any shortcomings.

sharp cuts might be made in the instructional budget, cuts that would have to be made without any measures of their educational impact. Together with a certain intellectual curiosity these events were sufficient to generate the project described in this paper—a major attempt to analyze the process of education in the introductory economics course.

This paper describes a major substantive study of economic education. The study included (1) construction of a three-equation model consisting of a "behavior function," an equation describing student decisions about the allocation of time and effort; a "production function" relating student effort, ability, and pedagogy to achievement in economics; and a "profit function" relating student effort and achievement to student enjoyment of the course; (2) collection over a three-year period of data on 2,400 students required to test the validity of that model; and (3) estimation of that model with regard for the simultaneity and nonlinearity inherent in the educational process.

The results of this project are interesting from several perspectives. They suggest, for example, that the maximizing model and standard technological assumptions of economics are useful for describing the educational decisions of college-age students. The results suggest that functions more general than Cobb-Douglas may be preferable when the data are sufficiently fine-grained. The elasticity of substitution can indeed differ from 1. But, what is most important, is that this study has produced a plethora of robust and interesting propositions about the educational process which speak to our original concerns.

1. Today we have a set of general propositions about the learning process in the course. For example:

- For the average student, the "elasticities" of ability, pedagogic inputs, and effort are roughly .89, .40, and .25;
- The largest influences on the time students devote to the course are tastes and outside commitments;
- Pedagogic inputs appear primarily to change the timing and mix of effort, rather than its absolute quantity.

2. An evaluation of our pedagogic innovations in detail reveals, for example:

- Self-paced instruction increases student achievement as measured by standard course exams, but not student enjoyment;
- Case-based instruction increases student enjoyment of the course but leaves achievement unchanged;
- Self-paced instruction increases the mean time spent on the course slightly, but sharply decreases the variance.

3. An analysis of the productivities of various educational inputs suggests that contrary to the conventional wisdom, for example:

- Neither age nor experience of instructor has a significant effect upon student performance;
- Student evaluation of instructors' competence, pedagogic skills, and empathetic qualities explains *none* of the variance in student performance on standard tests of understanding of economics;
- Some "objective" characteristics of instructors—grades in graduate courses and hours spent preparing classes—explain about 20 percent of the variance in student performance;
- Student evaluations of instructors' competence and pedagogic skills account for about 60 percent of the explained variance in student enjoyment of the course.

This listing of results is not intended to imply that the issues of economic education have been resolved by this analysis; indeed a more sophisticated mode of analysis invariably creates new problems even as it resolves old ones. Nevertheless, the results should be sufficient to persuade conventional economists that the application of a maximizing model to problems of educational choice is not just empty formalism, and to persuade economic educators that the costs of moving to a conceptually, empirically, and statistically more sophisticated method of evaluation are far outweighed by the value of the propositions uncovered.

The study is divided into five parts. Section I develops the model of educational choice. Section II describes the data and methods used to estimate the model. The results of estimating the model are presented in Section III. Section IV reports estimates for subgroups of special interest: very bright students, students of average ability, women

students, and self-paced students. Section V summarizes the findings, reservations, and unanswered questions.

1. THE MODEL

The model described below, while more elaborate than those usually found in the economic education literature, can be approached intuitively as follows: If one were asked to evaluate the results of an educational innovation or arrive at a judgment about the trade-off between intelligence and effort in an introductory economics course, an obvious first approach (and the one taken in most of the education literature)³ would be to estimate a model of the form:

$$[0] \text{ ACH} = a_0 + b_1 \text{ ABIL} + b_2 \text{ EFF} \\ + b_3 \text{ TEACH} + b_4 \text{ PED} + e$$

where

ACH = student mastery of economics

ABIL = student ability: Scholastic Aptitude Test scores, math background, etc.

EFF = measures of student effort: attendance, percent of homework completed, etc.

TEACH = instructor characteristics: experience, academic rank, preparation time, etc.

PED = innovation: 1 if participating student; 0 otherwise

and interpret the coefficients as marginal productivities of the various inputs. (A complete list of abbreviations with definitions is included at the end of this article.)

Upon reflection, however (particularly if one were familiar with the production literature), it would be clear that such a model provided rather limited information; a single-equation model necessarily provides information only on achievement, not on other outputs such as enjoyment or concentration. It would be equally clear that if, for example, an hour's extra study time does not produce the same gain in mastery for every student regardless of ability or if study is subject to diminishing returns (i.e., if the assumptions of linearity and additivity do not hold), or alternatively, if effort and

achievement are simultaneously determined, (i.e., if students work at what they're good at and are good at what they work at), then EFF and e would be correlated. If the assumption that $\sigma_{ee} = 0$ is violated, ordinary least squares (OLS) estimates will be biased and inconsistent; OLS estimates of b_2 , for example, will overstate the gains to the average student from increasing his efforts in the course.³

Fortunately, both problems are manageable. A natural first step is to deal with the simultaneity problem by the explicit introduction of an effort equation which recognizes the interdependence of effort and achievement.⁴ The education literature does not contain a model of the allocation of effort; however, one can be derived from the theory of utility maximization.

Assume that each student is a utility maximizer, endowed with some set of abilities and tastes (assumed to be predetermined) and a technology defined by the teaching technology and quality of the instructor assigned to him. The student controls one input, effort. For the time being, assume that effort is allocated along a single dimension and that this effort produces a grade and consumption benefits in some fixed relationship, that is, that one hour of work yields a fixed proportion of utility and marginal productivity with respect to grade received; no trade-off between the two outputs is possible. Time, a surrogate for effort, is allocated among three occupations: economics, all other courses, and all other activities. Utility is a function of the direct utility of the three time allocations and of the expected grade to be received in economics and other courses. (Thus we are in addition making an assumption of risk neutrality.) For an individual, let U_i = utility received by the i th student; T = student time; $\hat{G}$ = anticipated grade; EC = economics; AOC = all other courses; AOA = all other activities. Then

$$U_i = f(T_{EC}, T_{AOC}, T_{AOA}, G_{EC}, G_{AOC}) \\ T_{EC} + T_{AOC} + T_{AOA} = 24 \text{ hrs/day} \\ \hat{G}_{EC} = g(T_{EC}); \hat{G}_{AOC} = h(T_{AOC})$$

The marginal conditions are as follows:

$$(\partial U_i / \partial G_{EC}) (\partial G_{EC} / \partial T_{EC}) = (\partial U_i / \partial T_{AOA}) \\ - (\partial U_i / \partial T_{EC})$$

$$(\partial U_i / \partial G_{AOC}) (\partial G_{AOC} / \partial T_{AOC}) = (\partial U_i / \partial T_{AOC}) \\ - (\partial U_i / \partial T_{AOC})$$

Time is allocated to economics and to all other courses until the expected utility of the increased grade derived from the final hour of study minus the disutility of the study itself equals the (net) marginal utility of an hour spent in each other way. In particular, then, we would expect those with a taste for study to work more, and those with a taste for other activities to work less. Those with a taste for high grades would study more and those who believe themselves more efficient at the margin in studying (*ceteris paribus*) would study more.

Thus we have a behavioral function for the *i*th student which can be written as:

$$[1] \text{ Effort} = f_i([Ts], \hat{G})$$

where effort (an unobservable, to be approximated by homework hours, reading hours, attendance, etc.) depends on $[Ts]$, a vector of student tastes and interests, together with $\hat{G}$, the expected grade, an unobservable, assumed to be a function of past grades, current performance, and available technology, as defined by pedagogical inputs available to the student. In addition, "institutional" factors such as attendance requirements may enter directly.⁵ $[Ts]$ may be assumed exogenous to any one course. $\hat{G}$, to the extent it depends on current as well as past performance, is endogenous. Institutional constraints are, of course, exogenous.

Thus a first-semester effort function can be written as:

$$[1.1] \text{ EFF}^i = f([Ts]), [TEACH], \{G_{t-1}^{i-1} a_i G_{t-1}^{AOC}, \\ GR_{t-1}^{EC10}, ACH, [PED]\}$$

where

EFF^i = effort devoted to the introductory economics course by *i*th student in the first semester

$[Ts]$ = vector of tastes: intellectual, social, etc.

$[TEACH]$ = vector of instructor characteristics: empathy, classroom skills, preparation, etc.

G_t = previous grades, other courses

GR_{t-1}^{EC10} = previous grade in course

ACH = current performance in course

$[PED]$ = educational inputs other than instructor

It is not incompatible with a maximizing model to assume that in a two-semester course, study habits of the first semester influence choices in the second. The second-semester effort function then becomes⁶

$$[1.2] \text{ EFF}^i = f(\text{EFF}^i, [Ts], [TEACH], GR_{t-1}^{EC10}, ACH, [PED])$$

A natural second improvement is to relax the assumptions of linearity and additivity in the "production function," or achievement equation,⁷ by choosing a more general functional form.⁸ The CES specification, in particular, allows the effort-ability "isoquants," for example, to range from L-shaped curves to straight lines. Equation [0] then becomes:

$$[2.1] \text{ ACH}^i = A(\alpha_1 \text{ABIL}^{-\beta} + \alpha_2 \text{EFF}^{-\beta} + \\ \alpha_3 \text{TEACH}^{-\beta})^{-1/\beta} + \alpha_4 \text{PED}$$

where ACH^i = index of student mastery of economics: test scores and grades in the first semester; ABIL = vector of student abilities: general intelligence, analytic ability, mathematical training, etc.; and EFF , TEACH , and PED are defined as before. Given that economics is a cumulative discipline, the second-semester production function becomes:

$$[2.2] \text{ ACH}^i = A(\alpha_1 \text{ACH}_{t-1}^{-\beta} + \alpha_2 \text{ABIL}^{-\beta} \\ + \alpha_3 \text{EFF}^{-\beta} + \alpha_4 \text{TEACH}^{-\beta})^{-1/\beta} \\ + \alpha_5 \text{PED}$$

Finally, the model can be expanded to include what many consider to be an important product of an introductory course, namely, student satisfaction: not the answer to the question, "Are you having fun this minute?" but rather a more reflective answer to a question asked after the course, "Was it a worthwhile experience?" Such a satisfaction equation can be derived by returning to the maximizing model discussed earlier.⁹

Assume that students allocate their effort according to equation [1]. Given a student's achievement function, this effort (a cost)

results in some grade, $GR^{E:10}$ (observable), and some amount of intellectual satisfaction, SAT (an unobservable). By our earlier assumption, $GR^{E:10}$ and SAT are joint products.

Now we assume that, just as in the achievement equation students' effort was indexed by their ability, so, too, do they have different "happiness" technologies, defined by their talent for happiness, H (to be measured by general satisfaction with college life, etc.) and the educational inputs the course makes available to the individual student. Thus we can write:

$$U = H(SAT, GR^{E:10}) - g(EFF)$$

where $\partial U/\partial SAT > 0$; $\partial U/\partial GR^{E:10} > 0$; $\partial U/\partial EFF < 0$. We cannot observe SAT; however, if we assume that it depends on student tastes, we can rewrite the above as:

$$[3] \quad EN] = f[GR^{E:10}, [Ts], H, [PED]] - g(EFF)$$

Since there is no reason to believe that enjoying economics becomes a habit, the second-semester equation is identical to the first.

The model described by equations [1] through [3] represents one solution to the problems inherent in the single-equation linear model. By explicitly recognizing the simultaneities and nonlinearities inherent in education, and by extending the model to include students' decisions about the allocation of effort and judgments of satisfaction, it should be possible to arrive at both more accurate estimates of the marginal productivities of educational inputs and broader evaluations of the overall impact of any pedagogic innovation. Furthermore, since this model is merely a special case of the more general maximizing model, its usefulness in explaining student behavior will provide one more piece of evidence on the appropriate range of its application.

II. ESTIMATING THE MODEL

The model described by equations [1] through [3] can be estimated in three steps. First, it is convenient to linearize equations [2]; nonlinear estimation techniques exist but they are expensive and artful. Taking natural logs and making use of the Kmenta approximation (i.e., an expansion of the

Taylor series around $\beta = 0$)¹⁰ equation [2] can be written as:

$$\begin{aligned} \ln ACH = & \ln a + \alpha_1 \ln ABIL + \alpha_2 \ln EFF \\ & + \alpha_3 \ln TEACH + \alpha_4 SPI + \alpha_5 CMI \\ & - (1/2)\alpha_6 (\ln ABIL - \ln EFF)^2 \\ & + e_i \end{aligned}$$

where SPI = self-paced sections and CMI = case method sections. Intuitively, the squared term "corrects" for departure from Cobb-Douglas when $ABIL$ and EFF take on extreme values.

Second, having specified that ACH and EFF are simultaneously determined, an alternative to estimating equation [2] via ordinary least squares is required. Fortunately tastes and precollege career plans are highly correlated with effort but clearly exogenous to the course, as are Scholastic Aptitude Test scores and math preparation with achievement. Thus consistent (and reasonable) estimates of equations [1] and [2] are obtained by applying two-stage least squares, i.e., by estimating

$$\begin{aligned} EFF = & a_0 + a_1[ABIL] + a_2[Ts] + a_3[TEACH] \\ & + a_4[PED], \end{aligned}$$

replacing EFF by EFF in equation [2], and estimating equation [2] by OLS. This procedure is now legitimate since EFF is uncorrelated with the error term.

The third step is to choose operational equivalents for each of the terms in equations [1] through [3]. The body of available data, which included admissions data, data collected directly from students on "signed" questionnaires (with identification numbers), and information from course and college records, is very rich; conventions as to how many of the terms of the model should be measured are almost nonexistent. Thus experiments with data from spring 1973 were used to establish reasonable conventions for measuring tastes, effort, etc. These conventions were then adopted in 1974-1975 and 1975-1976. In spite of this experimentation the errors-in-variables problem remains important. Many of the observables that have the most explanatory power do not neatly correspond to one term of the model.

The data used are described briefly below.

1. Student tastes and interests [Ts]. Direct information on student tastes and interests was obtained primarily from a specially designed section of the course questionnaire, which asked for priority rankings among lists of activities with responses ranging from "agree-disagree" to statements such as, "I feel there are more important things to do than study." Factor analysis was then used to identify major dimensions of interests: academic (ACINT), personal (PERINT), entrepreneurial (PROF/ENTR), political (POLINT), and athletic (ATHINT), along which each student's interest could be described.¹¹ Students were also asked for information on their extracurricular activities.

Variables used as proxies for tastes included race (MTY), year in college (FROSH),¹² and sex. Variables used as proxies for the value of the expected grade included SECON (declared concentrator when entering course), PREMED, and a dummy P/F variable for students enrolled in the course on a graded/ungraded basis.

2. Instructor characteristics [TEACH]. Data were obtained on four "objective" instructor characteristics: intellectual ability, as measured by college and graduate school grades (TEACHGRADES); teaching experience (TEX); intellectual experience as measured by the overlap of academic specialization with course content (TFSPEC); and assignment policies, measured by the quality and quantity of problem sets and handouts provided (#PS). Student views of their instructors were summarized by two variables: a rating on the intellectual competence of the teachers' performance (TEACHCOMP) which was a weighted average of responses to the questions on "well-prepared?" "explains clearly?" "handles questions well?" and (TEACHNICE), an average of responses to ratings on their human qualities, "accessible?", "fair in grading?" and "cares about your learning?"

3. Other instructional inputs (PED). Other than instructors, the major systematic differences in instructional inputs are the result of designating several sections as self-paced (SPI) or case-method (CMI), the two ongoing educational experiments. Both are entered as dummy variables.

4. Grade: past (G_t , GR_{t-1}^{EC10}), current, expected. Data on students' previous grades are obtained from college records for: upperclass-

men and from admissions records for freshmen. Previous course grade (GR_{t-1}^{EC10}) is not treated as a continuous variable. Harvard grading is notoriously nonlinear. The distance between a B+ and an A- is much greater than between a B and B+; thus $\partial EFF / \partial GR_{t-1}^{EC10}$ is not necessarily monotonic. To allow for nonlinearities, past course grade was entered as a dummy variable: i.e., D_A^{EC10} , D_B^{EC10} , etc.

5. Student effort [EFF]. Measures of student effort included attendance, percent of reading done, percent of homework completed, hours of study time and, as reported by students on the questionnaire, "time spent on this course relative to other courses." These reports were cross-checked against course records to ensure consistency. Three proxies for effort were also used: SECON, P/F, and RPD, a dummy variable equal to 1 if a student was not in class to receive a questionnaire.

6. Student ability and preparation [ABIL]. Direct measures of student ability used are: Scholastic Aptitude Test scores (VSAT, MSAT), secondary school teachers' ratings of a student's academic promise (RATES1), and a Harvard admissions rating on intellectual promise (RATES2). Precollege plans to concentrate in the humanities (SHUMAN), high school math preparation (SUMS), race, and gender are used as proxies for the subtler components of ability.¹³ All ability data were obtained from admissions records under appropriate safeguards.

7. Achievement [ACH]. The central achievement measure (ACH) was the score on the multiple-choice portion of the semester final. This is a sixty-minute, forty-question security exam (i.e., no copies are allowed to leave the room), some portion of which is re-used. First-semester coverage includes supply and demand, price theory, labor, industrial organization and public finance. The second-semester exam covers macroeconomics, trade, and comparative systems. Questions from both the College Level Exam of Economic Understanding (CLEU) and the Test of Understanding in College Economics (TUCE) are included. Over the last three years the correlation between scores on the multiple-choice portion of the exam and the two-hour essay portion have ranged from .74 to .85.

8. Student satisfaction [EN], SAT]. Stu-

dent responses to the question, "How worthwhile did you find this course?" are the chief indicator of student enjoyment of the course (ENJ). Students' general happiness with college life was measured along two dimensions. One was intellectual involvement (INTINV), measured by responses to questions such as, "I would prefer all courses pass/fail." The other was satisfaction with college life (COLSAT), measured by responses to questions such as, "Harvard-Radcliffe has lived up to my expectations."

In fitting the model these data were used in two forms. An overview of the process is best achieved by using one overall index of each input or output. Thus for the "forest" version of the model, factor analysis was used to construct synthetic "ability," "effort" and "instructor" variables from the many partial indexes. The loadings are shown below (see list of abbreviations for definitions):

"Ability"	"Effort"	"Instructor" ¹⁴
VSAT	HOURS	TEACHGRADES
MSAT	ATTEND	TEACHEX
SUMS	% READING	TFSPEC
RATES 2	RELTIME	TEACHCOMP
RATES 1	RPD	
P/F		
SEX		
RACE		

This procedure, which extracts the "pure" component of each observable, also reduces the errors-in-variables problem created by observables such as P/F, year, and gender that do not correspond precisely to any one theoretical term in the model.¹⁵

Improving a course, however, requires knowledge not of the return to a metaphysical "effort" concept, but estimates of how much improvement an extra problem set will produce. To produce this managerial information, a "tree" version of the model, in which numerous partial indices of each theoretical term were entered as independent variables, was also estimated.

III. RESULTS

The two-stage least-squares estimates of the effort and achievement equations for 1975-1976 are presented in Tables 1, 2.1,

2.21, and 2.22. In addition to the "forest" and "tree" versions described above, equations [2] were also estimated in linear form to allow comparisons with other studies in economic education. The ordinary least squares estimates of the enjoyment equation are presented in Table 3. Results from earlier years are not presented; discrepancies, if any, are noted for each equation. The discussion of the results presented in each table is organized around the three central purposes of this exercise: the search for general propositions about the process of economic education; the examination of narrower pedagogic issues about the effects of educational experiments and instructor characteristics; and the resolution of the methodological questions of the usefulness of an economic approach to education.

A. Effort (Equations [1.1] and [1.2])

1. General propositions. The results presented in Table 1 suggest three propositions about the effort allocation process. The first is the importance of students' tastes (INTINV, PROF/ENTR, etc.), in their decisions as to how to allocate their time. Student interests, proxies for student interests, and involvement in extracurricular activities, including term-time employment, explain roughly twice as much of the variance in effort as any other group of variables.

The second proposition concerns the role of current performance and previous grades. The coefficient on previous grades, G_{t-1}^{AOC} , whether at Harvard or in secondary school, is robustly insignificant. The coefficients on previous course grades and on VSAT, MSAT, and SUMS (instruments for predicting current achievement) are consistently negative. Furthermore, the higher the recent or current grades, the lower the level of effort. The old maxim, "give them an A, and they'll stop working" seems to have some statistical validity. A comparison of results for the first and second semesters provides additional insight into the complex motivational role of course grades. Early (i.e., in October) a B- is a motivating grade: B- students work harder, ceteris paribus, presumably in hopes of becoming A- students. However, students who receive a B- at the end of the first semester work less during the second semester; it is possible that they perceive the nonlin-

TABLE 1 Effort; All Students

	First Semester		Second Semester	
	Coeff.	t-Ratios	Coeff.	t-Ratios
SATCOL	-.0416	(.081)	.0682	(.431)
INTINV	.6261	(3.80)	.3196	(.872)
PERINT	.1029	(1.800)	.1512	(1.239)
POLITINT	-.3315	(1.506)	-2.082	(.640)
PROF/ENTR	-.6769	(5.150)	-.1579	(.267)
ATHLETICS	.0473	(.013)	.1917	(.258)
WRITING	.8428	(2.802)	-.1693	(.184)
JOB	-.8347	(3.153)	.0082	(.000)
EXTRA	-.7598	(2.291)	.1042	(.072)
MTY	.0576	(.010)	-.2488	(.185)
FROSH	.5409	(2.173)	.2699	(.492)
SEX	-.2073	(.258)	.0432	(.008)
SECON	1.2938	(10.263)	.7708	(4.629)
P/F	-2.8279	(22.76)	-1.722	(3.731)
PREMED	.1165	(3.98)	.0776	(1.78)
TEACHCOMP	-.1889	(1.800)	-.0614	(.497)
TEACHNICE	-.1221	(.782)	.3217	(.007)
TEACHGRADES	-.0197	(.009)	.0297	(.084)
#PS	.1901	(.687)	-1.333	(1.377)
SPI	1.2976	(6.371)	.2987	(.384)
CMi	.1598	(.86)	.0017	(.081)
G ^{AOC} ₁₋₁	.0035	(.052)	.0163	(.837)
A	-.2270	(.244)	-.2884	(.364)
B-	.2888	(1.317)	-.7851	(2.964)
C-, C, C+	1.1989	(4.911)	-.1014	(.022)
D, E	.1808	(.119)	-.4184	(.313)
VSAT	-.0034	(1.954)	.0006	(.046)
MSAT	-.0078	(8.417)	.0041	(1.723)
SUMS	-.4299	(3.079)	-.5827	(4.888)
SMART	-1.1172	(2.083)	.1574	(.033)
EFF ¹	-	-	.6369	(103.154)
Constant	18.163		2.078	
R ²	.3547		.5027	
SEE	2.8656		2.619	

earity of Harvard grading and resign themselves to being on the wrong side of the great divide.

The third proposition is the importance of habit. The correlation between first- and second-semester effort is .64. Given the first-semester effort, estimates of the effort equation in the first-difference form reveal that only a change from graded to pass/fail status appears to have a substantial impact on student effort allocation.

2. Pedagogic issues. The results with respect to the pedagogical questions are mixed.

The coefficients on both problem sets assigned and SPI are positive, although the coefficient of problem sets is significant only when the variable is entered in log form. Examination of the distribution of effort scores among the self-paced group indicates that this difference is due primarily to the compression of the left-hand tail. Apparently past some point (in study time, about three hours per week), the demands of SPI or additional problem sets lead to substitution among varieties of effort; structured assignments drive out unstructured reading. The coefficient of the

case-method dummy is consistently insignificant. This result may simply be a reflection of the small number of cases assigned each term rather than any general proposition about the labor intensity of the case method.

The coefficients on all instructor attributes are generally small and insignificant. This is surprising. A student who draws a superior teacher can learn the same amount with less effort. This "income effect" might lead the student to work less. There is also a substitution effect, since the price of a given grade in terms of effort is lower in the course taught by the superior teacher. It seems implausible, however, that the substitution effect and the income effect should cancel precisely. Yet this result emerges in all three years.

3. *The maximizing model.* These results are generally encouraging with respect to the question of the appropriateness of a maximizing model. The taste coefficients have the "right" signs; would-be entrepreneurs, aspiring lawyers, and students with strong academic interests work more; those with strong personal concerns work less. The coefficient of *SECON* is large and positive; the coefficient of *P/F* is strongly negative; students planning a concentration in economics work more, while those taking the course pass/fail work less. Students, who have gotten an A (and thus in some sense "overinvested") correct their behavior and work less in the next round. Students who choose to take the course pass/fail during the second semester reduce their efforts on all dimensions: attendance, reading, and study time. Perhaps the major departure from what a pure maximizing model would have predicted is the apparent importance of first-semester habits during the second semester. This may, however, simply reflect the imperfections of the taste measures or stability in relative returns over the course of the year from other courses and other activities.

B. Achievement

1. *General propositions.* Of the results reported in Tables 2.1, 2.21, and 2.22, four are particularly informative.

a. *The importance of learning to learn.* In the beginning, i.e., during the first semester, the coefficients of ability (*ABIL*), student effort (*EFF'*), and instructor (*TEACH*) in Table 2.1 are, respectively, .84, .04, and .39. Interpreted as elasticities, these coefficients sug-

gest that a 1 percent increase in ability has roughly ten times the effect of an equal increase in effort. However, during the second semester the effort coefficient rises to .18: meaning that an extra effort unit (more study time, more reading, and more frequent class attendance) is more productive during the second semester than the first.

To some extent this increase may merely reflect the use of a better ability measure (i.e., first-semester achievement) rather than the students' acquisition of a better learning technology. However, the results estimating the detailed version of equation [2] (Tables 2.21 and 2.22) suggest that learning how to learn does occur. The pedagogic and instructional variables are uniformly less significant during the second semester. The coefficients of instructor grades and numbers of problem sets are insignificant and occasionally perverse. Self-paced instruction, which raises scores by about 15 percent in the first semester, has no significant effect in the second. When effort is entered in disaggregated form, the coefficients of the structured time variables—class attendance and problem sets—are large and significant during the first semester; the coefficient of *HRS* is very small: an extra hour of study time is worth .10 points on the exam. By the second semester, however, an extra hour of study time yields .27 points; *HRS* explains about 15 percent of the variance in achievement, while *ATTEND* and *#PS* become insignificant.

Obviously this finding is open to other interpretations. Nevertheless its implications concerning the scope for better pedagogy and for the allocation of instructional resources are profound.

b. *The multiple dimensions of ability.* When the composite ability variable (*ABIL*) is replaced by its components, *VSAT*, *MSAT*, *SUMS*, *SHUMAN*, etc., the R^2 of the equation (corrected for degrees of freedom) improves sharply. It appears that the "ability" relevant to achievement in economics is multidimensional, including, at a minimum, general verbal skills, mathematical aptitude, and mathematical preparation. Several clues suggest a fourth dimension. Even after formal mathematical preparation, mathematical aptitude, and effort are controlled for, students who intended to concentrate in humanities, women, and pass/fail students score lower. One can read into these results the existence of a fourth dimen-

TABLE 2.1 Achievement; "Forest" Version (Log Form)

	First Semester		Second Semester	
	Coeff.	t-Ratios	Coeff.	t-Ratios
ABIL	.8419	115.14	.8552	13.006
EFF ^a	.0401	2.249	.1764	8.068
TEACH	.3872	8.827	.0961	1.687
SPI	.0786	6.987	.0028	.002
ACH ^a	-	-	1.0896	79.872
(ABIL - EFF) ^a	.0402	.761	.0232	1.108
Constant	.0938		-3.4491	
R ²	.3351		.4326	
SEE	.20397		.4233	

TABLE 2.21 Achievement; All Students; "Tree" Version (Log Form)

	First Semester		Second Semester	
	Coeff.	t-Ratios	Coeff.	t-Ratios
VSAT	.5091	(40.11)	.6499	(7.378)
MSAT	.5422	(29.711)	.7026	(6.004)
SUMS	.0710	(7.415)	-.0316	(.135)
SHUMAN	-.1129	(13.068)	-.0856	(.929)
SEX	-.0073	(.112)	-.1073	(2.717)
MTY	.0197	(.376)	.0939	(1.087)
FROSH	.0167	(.696)	.0311	(.282)
HRS ^a (hours, 0-20)	.0201	(.867)	.1502	(2.132)
ATT ^a	-.0713	(11.961)	.0172	(.638)
RPD	.0644	(9.324)	.0421	(1.941)
TEACHCOMP	-.0073	(.153)	-.0251	(.227)
TEACHGRADES	.3374	(6.765)	.2017	(1.871)
#PS	.0756	(5.780)	.0155	(1.368)
P/F	-.1210	(2.938)	-.1909	(3.398)
SPI	.0787	(7.448)	.0089	(.012)
CMI	.0001	(.008)	.0012	(.000)
ACH ^a	-	-	1.0415	(60.622)
Constant	-4.633		-9.3006	
R ²	.3793		.4928	
SEE	.1939		.4029	

sion of ability: a talent for the peculiar model-building that is the heart of economics. If in addition one cared to postulate that students know something about their analytic ability, a puzzling result of the last section could be explained. Students who know themselves lacking in this analytic ability (as opposed to general intelligence) take the course pass/fail. It is then not surprising that differences in past performance not related to analytic ability, i.e., past grades in human-

ities courses, have no effect on the effort allocation decision.

c. *Returns to scale.* These results suggest that there are increasing returns to scale; that if ability, effort, and instructional quality are increased simultaneously, achievement will increase more than proportionately. This is somewhat surprising; it suggests the introductory course has scope enough to occupy even very bright students for a year.

d. *Substitution possibilities.* To some ex-

TABLE 2.22 Achievement; All Students; "Tree" Version (Linear Form)

	First Semester		Second Semester	
	Coeff.	t-Ratios	Coeff.	t-Ratios
VSAT	.0206	(48.509)	.0178	(16.536)
MSAT	.0241	(49.949)	.0218	(6.643)
SUMS	.8227	(8.701)	.0667	(.027)
SHUMAN	-2.5129	(12.718)	-.7753	(.593)
SEX	-.4058	(.668)	-2.0568	(8.443)
MTY	.6824	(.889)	1.353	(1.795)
FROSH	.4546	(1.005)	-.4079	(.328)
HRS ²	.0990	(.630)	.2706	(8.459)
ATT ²	-.7139	(11.22)	.2164	(.408)
RPD	1.3269	(7.725)	.5538	(.875)
TEACHCOMP	.0090	(.001)	-.0391	(.455)
TEACHNICE	-.0002	(.000)	.0021	(.002)
TEACHGRADES	.7769	(9.534)	.2430	(.455)
#PS	.4399	(11.640)	-.1459	(.514)
PIF	-3.5477	(3.468)	-1.6434	(2.31)
SPI	2.210	(7.510)	.4689	(.25)
CMI	.1188	(.731)	.0639	(.087)
ACH ¹	-	-	.8126	(166.2)
Constant	-15.4330		-26.873	
R ²	.4346		.5792	

tent knowledge of the "elasticity of substitution" among educational inputs is less helpful than in other processes; students cannot "move along" an ability-effort isoquant, nor can they recruit brighter instructors. Nevertheless, it is instructive to identify several such isoquants that show relative marginal productivities of ability and effort. They are relatively steep.

Questions of substitutability among other inputs are explored at length in Section IV where the assumption of constant elasticity is relaxed.

2. *Pedagogic issues.* The results of the pedagogical and instructional variables are somewhat limited by the assumptions of the model, in particular the assumption that $\partial^2 \text{ACH} / \partial \text{ABIL} \partial \text{TEACH} > 0$. Nevertheless, two pedagogic findings are among the most surprising of the entire study. The first is that instructional inputs—pedagogical methods and instructor characteristics—do matter. The coefficients of *SPI* and *TEACHGRADES* are consistently positive and significant. This result may imply that the production function approach is most valu-

able when the data set has been disaggregated to the point where problems of differing objective functions among instructors are less acute.

The second set of surprises concerns the winners and losers among instructional inputs. Self-paced instruction appears to have a significant effect on student learning during the first semester. *Ceteris paribus*, self-paced students have scores about 12 percent higher than students in conventional sections. Students using the case method do not score better than their counterparts in conventional sections.¹⁶ The self-paced coefficient is small during the second semester, probably reflecting both the fact that only six weeks of the second semester are self-paced and the increasing ability of students to learn economics on their own.

Similarly, the value added of a good instructor, particularly during the first semester, is significant. For a student of average ability who puts the average effort into the course, a switch from the worst to the best instructor has a predicted value of about 4 points on an exam with a mean score of 21, or the equivalent of 200 SAT points.

What matters about an instructor, however, are graduate school grades, assignments per student, and (in the single year for which such data are available) self-reported hours of preparation. These three characteristics have a significant correlation with student achievement. At the mean, for example, each additional point in the grade point average of the instructor translates into roughly an increase of .75 point in a student's score on the final exam. Student ratings of instructors on human qualities (*TEACHNICE*) or intellectual competence (*TEACHCOMP*) are uncorrelated with student achievement. This finding is not sensitive to the reweighting of ratings, or to the grade point average of students whose ratings are used. It is consistent over the three-year period. Neither instructor's experience (*TEX*) nor professional specialization is significant. The former is perhaps attributable to the narrow range of experience offered by this group, but it is somewhat surprising that instructors are not more effective the second time around.

These results do not suggest that instructors do not matter, or that students' reactions to their instructors are unimportant. The results of the enjoyment equation, for example, imply that the instructor is the largest single factor affecting a student's enjoyment of the course and choice of concentration. But these findings do suggest that student opinions about their instructors reflect factors other than their instructors' contributions to their learning when learning is measured by exam performance.

A third, less startling but equally important finding is implicit in the discussion of the previous section. When students begin their study of economics, there is very little they can do to help themselves. The marginal productivity of the input over which they exercise exclusive control—study hours—is very low. Given that in the short run their abilities are fixed (although over time they could increase their level of mathematical preparation), the only inputs that have a significant effect on their performance—attendance, the number and quality of problem sets assigned, and the ability of the instructor—are almost entirely out of their control in the typical course structure.

3. The maximizing model. At a general level the results presented in Tables

2.1–2.22 are consistent with a maximizing model of student learning. All the inputs have positive signs; their magnitudes and changes over time seem plausible. Slightly over half the variance in exam scores is explained in the most successful equation, which is probably close to an upper bound, given the reliability of the exam. Two findings, however, suggest that the neoclassical model is not entirely applicable. The coefficient of the correction term ($ABIL - EFF$)², while insignificant, was generally positive. This suggests that this process does not exhibit diminishing marginal returns, i.e., that the ability-effort isoquants might be straight lines. Section IV presents a set of regressions that explores this problem in greater detail. Second, the evidence that there is "technological progress" in learning economics raises the possibility that the estimated coefficients represent "average" productivities based partly on students who are not transforming effort and ability into achievement efficiently.

C. Enjoyment

1. General propositions. Estimates of the enjoyment equation are presented in Table 3. They suggest three observations. First, the typical student arrives at his evaluation of Economics 10 by a fairly straightforward process. He asks himself three questions: (1) "What grade did I get?" (2) "Did my instructor run a good class?" (3) "Was my instructor nice?" The three variables that answer these questions: GRE_{10} , *TEACHCOMP*, and *TEACHNICE* account for 70 percent of the explained variance in students' retrospective enjoyment of the course. Among those three variables the enjoyment elasticities of *TEACHNICE* are slightly higher than teachers' grades.

Second, as in the effort equation, student tastes count, explaining about 25 percent of the variance. Students interested in entrepreneurship enjoy the course more and those interested in "people" enjoy the course less. Variables that are proxies for tastes—gender, race, and intended concentration—are also significant. Tastes appear to develop over time. The second-semester coefficient of gender is large and negative; women appear to reflect on the course less fondly over time. Conversely, minority-group students like it better. The decrease in the size of the

TABLE 3 Enjoyment; All Students (Log Form)
(range, 1-7; 1 = high)

	First Semester		Second Semester	
	Coeff.	t-Ratios	Coeff.	t-Ratios
SATCOL	-.1374	(1.403)	-.2953	(5.56)
INTINV	.2020	(1.617)	.0107	(.004)
PERINT	.1071	(.313)	-.0179	(.008)
ATHINT	.0202	(.034)	-.0797	(.384)
POLITINT	-.0801	(.415)	.0773	(.471)
PROF/ENTR	.3328	(5.563)	.3566	(6.371)
MTY	-.3890	(2.195)	-.5527	(4.42)
SEX	-.0186	(.010)	-.6183	(8.909)
SECON	-.7057	(12.89)	-.2460	(1.715)
TEACHEX	.0014	(.119)	-.0021	(.287)
TEACHCOMP	.0054	(8.003)	.0698	(8.751)
TEACHNICE	.1142	(11.261)	.0981	(5.44)
TEACHGRADES	-.0529	(.287)	-.0563	(.239)
#PS	.0327	(.347)	.1351	(5.077)
SPI	-.1511	(.395)	-.1281	(.309)
CMI	-.4117	(2.891)	-.2641	(2.319)
PIF	-.0938	(.113)	-.1325	(.169)
EFF	.0236	(.917)	-.0685	(7.72)
ACH	-.0074	(.135)	.0063	(.187)
SEMESTER	-.0799	(2.94)	-.1654	(13.41)
GRD (0-15)				
NOV GRADE (0-15)	-.0913	(5.840)	-	-
Constant	3.5100		4.725	
R ²	.3509		.3661	
SEE	1.2503		1.3241	

SECON coefficient in the second semester may reflect a taste for the more analytic micro material covered in the first term.

Exam score (ACH) has no independent effect on enjoyment; apparently the major source of satisfaction is not in knowing, but in having one's instructor know that one knows.

2. Pedagogic questions. The most interesting results about instructional inputs emerge from a comparison of Tables 2 and 3. Precisely those variables, *TEACHNICE* and *TEACHCOMP*, which are consistently insignificant in the achievement equation have large and significant coefficients in the enjoyment equation. Conversely *TEACHGRADES* and *#PS* are insignificant in the enjoyment equation, but significant in the achievement equation. Both self-paced and case-method instruction exhibit a similar pattern. *SPI* increases achievement but not

enjoyment even after controlling for additional effort; the converse is true for *CMI*. It is interesting that while still significant, the coefficient of *CMI* is about one-third its size in the first year of the experiment.

These results do not imply that there is a trade-off between student achievement and enjoyment. Rather, enjoyment and achievement are simply "produced" with different instructional resources. If additional pedagogic inputs are costly, careful thought should be given to the achievement-enjoyment package to be produced.

3. The maximizing model. The implication of these results for the appropriateness of the economic model is unclear. The taste and achievement variables have the expected signs. Contrary to our original hypothesis, however, there appears to be no significant negative relationship, *ceteris paribus*, between the effort a student reports investing

in the course and his enjoyment of the course. In the first semester the relationship, although negative, is insignificant; the second-semester coefficient although small, is perverse. (However, when the equation is estimated in first-difference form, i.e., when ΔENJ is the dependent variable, EFF has the correct sign.) It may be that for some students enjoyment is the "plug" factor; given his allocation of time, a student in retrospect adjusts his evaluation so as to make it rational. Alternatively, the assumption that an hour's effort produces identical grade-enjoyment bundles for all students may be inappropriate.¹⁷

IV. SUBREGRESSIONS

This section presents estimates of the model for three pairs of subpopulations: students of above- vs. below-average ability, men vs. women, and self-paced vs. conventional-format students. It has two purposes. The first is to resolve some of the issues raised in the previous section, particularly those of the substitution possibilities and scale factors among inputs. The second is to demonstrate the power of this model by attacking two questions that have been explored at length in the economic education literature with the standard single-equation model: (1) the efficacy of self-paced instruction and (2) the consistently inferior performance of female students in introductory economics courses.

A. Unresolved Issues

The results discussed in Section III left unresolved three questions with major implications for the allocation of instructional resources. One is the elasticity of substitution between ability and effort, where attempts to test the hypothesis that it differed significantly from 1 were inconclusive. A second was the appropriateness of the assumption of positive cross products: that the marginal product of better instructors was higher for bright students than for duller ones.¹⁸ The third was the finding that this process was characterized by increasing returns to scale, a finding that runs counter to both the standard economic assumptions and much of the education literature.

Fortunately, each of these questions is empirically testable. If at extreme values the elasticity of substitution between effort and

ability decreases, the coefficients of $ABIL$ should be lower for the brighter student, and those of EFF lower for the less able student. If the assumption that $(\partial^2 ACH / \partial TEACH \partial ABIL) > 0$ is appropriate, the coefficients of the instructor variables ($TEACHGRADES$, $\#PS$, etc.) should be larger the brighter the student. Finally, if the process is characterized by increasing returns to scale, the coefficients on all variables must be larger for the most able than for the least able. Thus each of these issues can be investigated by re-estimating the model for the top and bottom quarters: roughly those with MSAT scores above 725 and below 650.

Those results are presented in Tables 4.1, 4.21 and 4.22. Results for the instructor characteristics included in this analysis are surprisingly (and somewhat sadly) supportive of the assumption of positive cross products. In particular, brighter instructors seem to be much more useful to brighter students. The coefficient of $TEACHGRADES$ is positive and significant for the top quarter. $TEACHGRADES$ is negative but not significant for the bottom quarter. Similarly, problem sets seem to be more useful to the brighter student. In contrast the SPI format appears to be considerably more productive for the bottom quarter. The coefficient of SPI for the high group is not significant; for the bottom quarter it appears to add about 3 points to their exam score, or the equivalent of 150 points in SAT scores.

The regressions provide no strong evidence of increasing returns to scale over the entire range. The sum of the coefficients for the brightest and the least bright students is slightly below 1, suggesting a region of constant returns.

The issue of substitution is more complicated. The coefficient of $ABIL$ is lower for the brightest students, as predicted, although comparisons of the coefficients of $SUMS$ and $VSAT$ suggest that some components of intelligence may not be subject to diminishing returns. The coefficients of EFF and HRS , however, are larger for the bottom quarter.

Finally, as a by-product these results shed some light on two puzzles. The first is the finding that instructors have little effect on the allocation of student effort. At the lower end of the distribution, objective instructional quality appears to have a negative effect on effort. The bottom quarter is sen-

TABLE 4.1 Effort; Students Grouped by Ability
(top = MSAT above 725; bottom = MSAT below 650)

	Top Quarter		Bottom Quarter	
	Coeff.	t-Ratios	Coeff.	t-Ratios
INTINV	0.4089	(0.425)	0.557	(0.686)
PERINT	1.2035	(2.209)	-2.2716	(3.860)
POLITINT	-0.5839	(1.324)	-0.4896	(0.716)
PROF/ENTR	0.2146	(0.148)	-0.24751	(15.458)
ATHLETICS	0.0754	(0.009)	0.2841	(0.101)
WRITING	1.7723	(3.317)	0.9484	(0.766)
JOB	-0.9811	(1.150)	-0.1797	(0.029)
EXTRA	-1.0671	(1.474)	-1.2595	(1.233)
MTY	-1.6488	(1.451)	0.6806	(0.578)
FROSH	0.5665	(0.647)	1.1978	(2.213)
SEX	-0.0516	(0.004)	0.0297	(0.001)
SECON	1.3364	(2.777)	0.5539	(0.390)
P/F	-2.4501	(4.035)	-6.2830	(21.259)
PREMED	0.6124	(4.011)	0.4531	(1.887)
TEACHCOMP	-0.0243	(0.074)	-0.5713	(5.678)
TEACHNICE	-0.1121	(0.732)	-0.2103	(0.832)
TEACHGRADES	0.0772	(0.038)	-0.9199	(4.076)
#PS	0.2581	(1.735)	-0.4337	(3.930)
SPI	1.0404	(0.984)	2.7833	(12.654)
CMI	0.3307	(1.239)	0.0153	(0.019)
G_{t-1}^{AOC}	0.0054	(0.040)		
A	0.3154	(1.154)	4.4508	(8.101)
B-	1.0814	(1.842)	-2.2009	(8.346)
C	0.0756	(0.005)	-1.3649	(2.321)
D, E	-1.1280	(0.037)	-1.6168	(5.346)
VSAT	-0.0196	(1.471)	0.0978	(4.563)
MSAT	-0.0417	(3.472)	0.1099	(7.270)
SUMS	-1.0275	(4.310)	2.8256	(15.300)
Constant	31.069		38.4282	
R ²	.3835		.7100	
SEE	2.92773		1.9573	

sitive to *TEACHGRADES* (the higher the grade, the less the effort) and number of problem sets (more problem sets translate into less effort): a class geared too high is apparently discouraging. The coefficient of *TEACHCOMP*, however, is large and significant; the bottom group works much harder when an instructor is perceived as pedagogically competent.

The second is the perverse sign on *EFF* in the enjoyment equation. For the top quarter, *EFF* and *HRS* are consistently insignificant; it may be that these students are sufficiently talented so that over a low minimum, time spent on the course is consumption.

B. Some Standard Problems: A Re-analysis

1. *Self-paced instruction.* Evaluating an innovation is the most common occasion for an article in economic education. As noted in Section I, the heart of these evaluations is usually the estimation of a single-equation model:

$$ACH = a + b_1(MSAT + VSAT) + b_2SEX + b_3YEAR + b_4M \quad [1 = \text{student had calculus}] + b_5[1 = \text{the student trying new pedagogy}] + e$$

TABLE 4.21 Achievement; "Forest" Version; Top- vs. Bottom-Quarter Students
(figures in parentheses below coefficients are *t*-ratios)

	First Semester		Second Semester	
	Top Quar.	Bot. Quar.	Top Quar.	Bot. Quar.
ABIL	.6335 (10.58)	.785 (1.308)	.8606 (3.73)	.9368 (3.688)
EFF	-.0378 (.519)	.0669 (.92)	.1258 (1.619)	.2259 (1.38)
TEACH	.3412 (6.135)	.1613 (3.187)	.1789 (.081)	-
SPI	.0617 (1.087)	.1165 (3.927)	.0862 (.444)	.0334 (.069)
ACH ^a	-	-	1.0879 (44.864)	1.2586 (26.98)
Constant	1.1564	1.0682	-3.5508	-4.733
R ²	.1114	.2168	.3358	.4773
SEE	.2352	.2126	.4441	.4075

to yield a gross estimate of the effect of the innovation on achievement.

The analysis of SPI already presented moved past the analysis of an innovation conventionally presented in several respects. Explicit recognition of effort as an endogenous variable permitted unbiased estimates of the SPI effect. Introduction of the effort equation allowed the identification of a "pure" self-paced effect, i.e., its effect net of increases in effort. Estimates for the top and bottom quarter revealed something about the distribution of the gains.

The Ec 10 model, however, permits the investigation of how SPI improves performance to be carried one step further. More specifically if the model is re-estimated for SPI and for conventional format students separately and the coefficients of the effort, ability, and instructor variables are compared, it is possible to discover how SPI changes the educational production process. Some of the results of doing so are presented in Tables 5.1, 5.2, and 5.3.

Vis-à-vis the conventional group the most significant differences are as follows. (1) The coefficient of *ATTEND* (class attendance) is consistently significant for the conventional group, but larger and significant for the self-paced group. (2) The returns to mathematical preparation (*SUMS*) are only about half as large for the self-paced group as for the conventional group. (3) Instructor variables *TEACHGRADES* and *#PS* are significant for the conventional group, but not for the self-

paced group; they add nothing to the predictive power of the SPI achievement equation. (4) The traditional gender disadvantage disappears; the performance of self-paced women is not significantly different from that of men.¹⁹ (5) In the effort equation, current performance is negatively related to effort for students in conventional sections; it is insignificant for self-paced students. (6) Most important, the first-semester coefficient of *HRS* (work out of class) for self-paced students is as large as the corresponding second-semester coefficient for conventional students, or triple the first-semester coefficient for conventional students.

Taken as a whole these results suggest several hypotheses about how SPI works. Its major contribution appears to flow from its effect on student effort. It is somewhat imperialistic in that it encourages students to spend time on the course regardless of their current level of performance. At the same time, by providing the student with a highly structured learning environment and well-defined study strategies, it raises the productivity of a student's own time and effort. In the vocal idyll of production theory, this is a labor-saving rather than a capital-saving technology. With this labor-enhancing effect, self-paced students choose to spend more time on the course, which is reassuring, as it implies that achievement in economics is not an inferior good.

At the same time, these results suggest the limits of instructional inputs. For students in

TABLE 4.22 Achievement; "Tree" Version; Top- vs. Bottom-Quarter Students (Linear form)

	First Semester				Second Semester			
	Top Quar.		Bottom Quar.		Top Quar.		Bottom Quar.	
	Coeff.	t-Ratios	Coeff.	t-Ratios	Coeff.	t-Ratios	Coeff.	t-Ratios
VSAT	.0238	(16.68)	.0195	(9.809)	.0155	(3.188)	.0176	(4.327)
MSAT	.0219	(1.867)	.0084	(.617) ^a	.0362	(2.24)	.0209	(2.369)
SUMS	1.4833	(8.894)	1.081	(3.286)	.1297	(.031)	-.5421	(.490)
SECON	-.1027	(.012)	-.6479	(.326)	-.0112	(.007)	.3695	(.062)
SHUMAN	-3.3429	(5.507)	-1.691	(1.332)	-1.1163	(.343)	.3229	(.027)
SEX	.0093	(.000)	-.1966	(.036)	-1.3585	(1.159)	-2.984	(4.192)
MTY	2.1129	(1.734)	.0769	(.005)	1.1198	(2.48)	.8065	(.309)
FROSH	.8848	(1.001)	1.1367	(1.265)	-2.1320	(.033)	-1.289	(.938)
HRS ¹	-.1280	(.943)	.0866	(.283)	.2525	(2.203)	.4788	(4.609)
ATT	-.8199	(5.725)	-.8042	(2.321)	-.0361	(.004)	.1836	(.060)
RPD	-1.7319	(4.067)	-1.1935	(1.553)	.5348	(.253)	.3871	(.101)
TEACHCOMP	-.0336	(.209)	.0756	(.525)	-.0849	(.447)	.0943	(.574)
TEACHGRADES	.6479	(1.997)	-.0690	(.013)	.3705	(.334)	-.1864	(.060)
#PS	.5353	(5.157)	.1097	(.145)	-.1617	(.168)	.1652	(.169)
P/F	-2.4062	(3.202)	-.6089	(.125)	-.8760	(.223)	-1.798	(4.609)
SPI	1.225	(1.001)	2.5809	(3.607)	1.2961	(.586)	-1.080	(.004)
ACH ¹	-	-	-	-	.9160	(82.06)	.8057	(40.81)
Constant	-14.336		7.2966		-47.44		-28.4461	
R ²	.3229		.2884		.554		.5370	
SEE	4.7973		4.7269		5.7526		5.316	

^aMost students in this group have scores between 630 and 650; hence, the variance of MSAT is low.

**TABLE 5.1 First-Semester Effort;
Self-paced Students Only**

	Coeff.	t-Ratios
INTINV	1.929	(2.258)
PERINT	1.3342	(.845)
POLITINT	-1.9145	(2.979)
PROF/ENTR	1.1016	(.571)
ATHLETICS	-.6772	(.217)
WRITING	1.4954	(.834)
JOB	-.7188	(.195)
EXTRA	-1.4863	(.863)
MTY	-.0338	(.02)
FROSH	-.7131	(.277)
SEX	-1.4655	(1.061)
SECON	1.5815	(1.639)
P/F	-4.5584	(.589)
PREMED	.1137	(.726)
TEACHCOMP	-.0024	(.000)
TEACHNICE	.1121	(.004)
TEACHGRADES	-.0710	(.043)
#PS	.1146	(.008)
G ^{AOC} _{t-1}	.0028	(.167)
A	.7650	(.233)
B-	1.2123	(.535)
C	2.2849	(1.297)
D, E	2.4378	(1.253)
ACH ^d	-1.225 ^a	(7.418)
Constant	18.5154	
R ²	.1071	
SEE	3.5689	

^aACH^d is the regression of VSAT, MSAT, SUMS, and SMART on ACH.

a self-paced section the instructor's general intelligence, preparation, and assignments are uncorrelated with achievement. To the extent that assignments, for example, resemble unsupervised SPI, these findings reinforce the suggestion that as a learning device, problem solving is subject to diminishing marginal returns.

2. Explaining differences in male-female achievement. The question of why men consistently outperform women on standardized tests of economic achievement has been discussed at great length in the economic education literature. The range of explanations includes fathers who talk about business and finance to their sons but not their daughters, inadequate mathematical preparation, and fears of success as defeminizing.²⁰ Thus to ask that this model offer new

**TABLE 5.2 Achievement; First-Semester
Self-paced Students Only;
"Forest" and "Tree" Versions**

	Coeff.	t-Ratios
"Forest" Version		
ABIL	.8168	(29.114)
EFF ^f	.1522	(1.677)
INST	.0128	(.337)
Constant		.6282
R ²		.3178
SEE		.1433
"Tree" Version		
VSAT	.0127	(3.170)
MSAT	.0261	(9.745)
SUMS	.4503	(4.320)
SECON	-.0466	(.022)
SHUMAN	-2.0422	(1.186)
SEX	-.1128	(.123)
MTY	.3405	(.034)
FROSH	1.9550	(2.646)
HRS ^f	.2527	(2.459)
ATT ^f	-.8309	(2.043)
RPD	-.1241	(.011)
TEACHCOMP	-.0222	(.039)
TEACHGRADES	.5561	(.958)
#PS	-.4224	(.261)
P/F	1.9649	(.109)
Constant	-5.3515	
R ²	.4445	
SEE	3.8584	

insights on the gender-related differential is a fair test of its usefulness.

The results presented in Tables 2.1 and 2.2 provided one insight into this differential, namely, that the gap appears to be cumulative. In November, men and women perform equally well, in January, men perform slightly better, and by June they perform substantially better. Moreover, the insignificant coefficient on gender for both semesters of the effort equation indicated this cumulative gap was not attributable to a reduction in effort by women students.

When equations [1] and [2] are estimated separately for men and women, however, an even more interesting result emerges (Tables 6.1-6.3—pp. 190-93). The most striking difference between the male and female results is in the effort coefficients. The marginal product of HRS, an additional

**TABLE 5.3 First-Semester Enjoyment;
Self-Paced Students Only**
(range 1-7: 1 = high)

	Coeff.	t-Ratios
SATCOL	-.6165	(2.465)
INTINV	.2937	(.298)
PERINT	-.7589	(1.697)
ATHINT	.3451	(.767)
POLITINT	.5018	(1.243)
PROF/ENTR	.8206	(2.274)
MTY	-.8476	(1.824)
SEX	.6675	(1.302)
FROSH	.3196	(.346)
SECON	-.7967	(2.139)
TEACHX	-.6332	(.686)
TEACHCOMP	-.0625	(.504)
TEACHNICE	.1905	(3.502)
TEACHGRADES	-.0536	(.024)
#PS	-.2968	(.342)
P/F	.9903	(.151)
EFF	.0545	(.618)
ACH	-.0742	(.981)
SEMESTER GRD	.0618	(.244)
Constant	9.9578	
R ²	.1128	
SEE	1.482	

hour spent by a woman on the course, is less than half the marginal product of an additional hour spent by a man. Furthermore, during the second semester, when male "marginal effort product" (i.e., the coefficient of *HRS*) increases sharply in both size and significance, female marginal effort product remains almost zero: there is no significant difference between the first- and second-semester coefficients for women.

These results suggest the following account of the source of differential performance. Women enter the course with less "skill" in learning analytic material—less practice in that peculiar intellectual exercise of model building.²¹ Thus learning analytic material comes slowly; much of each hour spent studying is misdirected. Consequently at the end of the semester they have learned relatively little economics per unit of time and even less economic theory. Nor have they learned on the average to master analytic material as well as their male counterparts.

Thus in the second semester women are,

given the cumulative nature of economics, doubly disadvantaged. Since as the size of the coefficient of $GR^{E_{10}}$ in Table 6.1 indicates, they are more sensitive to grades than their male counterparts, they do not reduce their effort in the second semester. But as indicated by the large negative coefficient on gender in the second-semester enjoyment equation, they ultimately find the experience relatively unsatisfying.

This account is supported by three additional pieces of evidence. The first is a set of regressions, not reported here, in which the dependent variable of equation [2] is the subscore on theory questions only. In these regressions the gender differential for the first semester is approximately 15 percent and significant at the 1 percent level, suggesting that the problem does lie in the analytic area. The second, noted earlier, is the insignificance of the gender differential in self-paced sections, where drill in analytic materials is most intense. The third is another series of subregressions in which equation [3] is estimated for the hard-core humanities (*SHUMAN*) concentrators (those who planned on a humanities concentration before entering). The pattern of first- and second-semester effort coefficients for this group is very similar to the women's pattern while the gender coefficient is insignificant.

V. CONCLUSIONS

Traditionally a conclusion serves several purposes. It summarizes the findings, sounds the major themes, reviews the principal caveats and suggests directions for future research. The summary and major themes have already been provided in the introduction; therefore these concluding remarks deal with caveats and extensions.

The first set of caveats concerns the dependent variable. All of the analysis presented here deals with the determinants of short-run mastery of economics. While it seems plausible that the determinants of short-run mastery and those of long-run retention are similar, it is a hypothesis that remains to be tested. If they are not, then this analysis is much less interesting. In addition, even as short-run analysis, the dependent variable measures one dimension of mastery. While the correlation between multiple-choice scores and scores on essay questions is reassuringly

TABLE 6.1 First Semester Effort; Male and Female Students

	Male		Female	
	Coeff.	t-Ratios	Coeff.	t-Ratios
INTINV	.0618	(.028)	1.9282	(1.834)
PERINT	.7260	(2.210)	-.9880	(.305)
POLITINT	-.6392	(3.893)	-.3192	(.137)
PROF/ENTR.	-.8356	(6.019)	-.2264	(.021)
ATHLETICS	.2984	(.400)	-.6904	(.129)
WRITING	1.2414	(4.702)	-1.1373	(.230)
JOB	-1.1041	(4.148)	.9706	(.225)
EXTRA	-.5030	(.765)	-.4081	(.031)
MTY	.0924	(.019)	-	-
FROSH	.8266	(3.902)	.7389	(.215)
SECON	1.583	(11.410)	.1261	(.005)
P/F	-3.4030	(20.719)	-5.3755	(3.447)
PREMED	.1238	(1.871)	-	-
TEACHCOMP	-.0715	(3.107)	-.1566	(1.260)
TEACHNICE	-.0014	(.002)	-.0214	(.073)
TEACHGRADES	.6310	(1.88)	.1828	(.045)
#PS	-.1395	(.344)	-.2630	(.320)
SPI	1.3560	(5.149)	1.991	(1.874)
CMI	.1602	(.172)	-	-
A	.3395	(.479)	.9259	(.173)
B-	.0470	(.006)	.1410	(.005)
C-, C, C+	1.2009	(3.436)	1.3574	(.418)
D, E	-.3842	(.595)	-	-
ACH ¹	.5421	(13.096)	-1.0729	(2.370)
Constant	20.4838		17.4759	
R ²	.3196		.197	
SEE	2.7416		4.1097	

high, there is no direct evidence for the relationship between exam scores and the problem-solving skills that some instructors have made the focus of an introductory course. Again, while it would be surprising if students who did badly on standard material did well at problem solving, it is a relationship that needs to be investigated.²² In some sense, these questions about the dependent variable are the special case of the larger problem of measurement. Comparatively little is known about appropriate measures of effort, analytic ability, and tastes. It is therefore difficult to be confident that one has dealt satisfactorily with the errors-in-variables problem.

The principal questions and extensions suggested by this work are four. The first and most obvious is to test another group of introductory economics students. In par-

ticular it would be interesting to find a school where the upper quarter had MSAT scores around 650 and discover whether our "bottom-quarter findings" were a relative or absolute phenomenon. Second, as noted above, these results hold for a standard introductory course. It would be enlightening to compare these coefficients with those obtained by estimating the same model for an institutionally oriented or problem-based course. Third, because there has been no substantial change in curriculum or course structure (apart from the SPI and CMI experiments), this study has provided no data on the productivity of "nonteacher" inputs. It would be informative to re-estimate this model in a situation where there was a greater variety of format (some large-lecture, some small-section), changes in texts, or changes in depth or breadth of the

TABLE 6.2 Achievement; Male and Female Students; "Forest" and "Tree" Versions

	First Semester				Second Semester			
	Male		Female		Male		Female	
	Coeff.	t-Ratios	Coeff.	t-Ratios	Coeff.	t-Ratios	Coeff.	t-Ratios
Forest version^a								
ABIL	.927	(124.249)	.7887	(31.238)	1.8559	(80.374)	1.341	(6.631)
EFF ^b	.0431	(2.420)	.0234	(.102)	.2712	(14.012)	-.2287	(2.036)
INST	.3794	(8.662)	.1785	(.573)	.0498	(1.116)	1.4912	(2.062)
SPI	.0793	(5.642)	.0486	(.930)	.1202	(1.857)	.1129	(.400)
Constant	-.1259		.6492		-2.6526		-4.2859	
R ^a	.3053		.3223		.2778		.3142	
SEE	.2123		.1689		.4721		.4783	
"Tree" Version								
VSAT	.0218	(40.832)	.01601	(5.897)	.0173	(13.101)	.0174	(2.381)
MSAT	.0236	(34.508)	.0280	(14.982)	.0160	(3.765)	.0179	(1.992)
SUMS	.8167	(6.038)	.5992	(1.048)	.4153	(3.765)	-1.1585	(1.992)
SHUMAN	-3.2876	(13.009)	-1.0098	(.830)	-.6437	(.270)	-1.1345	(.346)
MTY	.5697	(.453)	1.2113	(.663)	-	-	-	-
FROSH	.5389	(1.097)	.4252	(.211)	-.2715	(.154)	.2833	(.028)
HRS ^c	.0309	(1.0846)	.0099	(.005)	.1810	(12.436)	-.0068	(.001)
ATT ^d	-.5464	(4.268)	-.7560	(3.907)	-.0561	(.063)	-.0916	(1.53)
RPD	1.2224	(4.873)	.5083	(.265)	.8099	(1.491)	-.3218	(.047)
TEACHCOMP	-.0137	(.078)	.0226	(.071)	-.0858	(.615)	.0984	(.288)
TEACHGRADES	.7988	(7.511)	.5434	(.988)	.1779	(.207)	.8389	(.678)
#PS	.4401	(8.989)	.4393	(2.188)	-.2335	(1.047)	.1807	(.119)
PIF	-1.3723	(2.080)	-2.1362	(2.208)	-1.927	(.865)	-3.7218	(2.409)
SPI	1.8583	(6.463)	1.1394	(1.646)	.4312	(.176)	1.4828	(.409)
ACH ^e	-	-	-	-	.7744	(128.025)	.8473	(21.745)
Constant	-17.1705		-12.4916		-23.785		-38.395	
R ^a	.4461		.4024		.5796		.557	
SEE	4.4942		4.2198		5.4662		5.807	

TABLE 6.3 Enjoyment; Male and Female Students
(range, 1-7; 1 = high)

	Male		Female	
	Coeff.	t-Ratios	Coeff.	t-Ratios
SATCOL	-.0389	(.093)	-.7564	(4.080)
INTINV	-.2282	(1.71)	.0074	(.001)
PERINT	-.0690	(.104)	.6138	(1.053)
ATHINT	-.0018	(.004)	-.191	(.390)
POLITINT	-.0874	(.373)	.0727	(.038)
PROF/ENTR	.3894	(6.783)	-.4559	(.329)
MTY	-.34161	(1.384)	-.4559	(.329)
FROSH	.0921	(.261)	.1515	(.083)
SECON	-.3359	(2.577)	-.8776	(2.953)
TEACHEX	.2303	(1.335)	.2545	(.187)
TEACHCOMP	.0664	(8.320)	.0586	(.919)
TEACHNICE	.1306	(11.617)	.0688	(.407)
TEACHGRADES	-.0305	(.087)	-.0001	(.016)
#PS	-.0018	(.001)	.1019	(.389)
SPI	.0111	(.002)	-.0695	(.008)
P/F	-.1481	(1.205)	.0764	(.119)
EFF	-.3078	(1.736)	.0123	(.068)
ACH	-.0085	(.142)	-.0361	(.033)
SEMESTER GRD	-.1247	(7.253)	-.1935	(2.674)
Constant	2.5495		5.6207	
R ²	.3509		.1121	
SEE	1.2375		1.648	

syllabus. It would be useful to compare these results with those in a course where macro was taught first. Finally, as the R^2 suggests, there is still a considerable amount to be learned about the determinants of student effort and enjoyment.

LIST OF ABBREVIATIONS

ABIL = student ability
 ACH = student mastery of economics (0-40)
 ACH' = achievement in first semester
 ACH' = fitted estimate of ACH
 ACINT = academic interests (1 = very important; 4 = not at all)
 AOA = all other activities
 AOC = all other courses
 ATHINT = athletic interests (1 = very important; 4 = not at all)
 ATT or ATTEND = attendance (1-7; 1 = always attends)
 ATT' = fitted estimate of ATT
 CLEU = College Level Exam of Economic Understanding

CMI = case-method sections (1 = yes; 0 = no)

COLSAT = satisfaction with college life

D_{Ec10}^{Ec10} , etc. = dummy variable for past grade in introductory economics

e = error term

EC = economics

Ec10 = introductory economics course

EFF = student effort (0-35)

EFF' = fitted estimate of EFF using exogenous variables only

ENJ = student satisfaction with the course

FROSH = freshman year of college (1 = freshman; 0 = other)

$\hat{G}$ = expected grade

G_p = previous grades in other courses

GR_{t-1}^{Ec10} = previous grade in introductory economics

H = happiness

HRS = hours of study (0-20)

HRS' = fitted estimate of hours of study

INTINV = Intellectual involvement (1 = not at all; 4 = very much)

MSAT = score on mathematics portion of Scholastic Aptitude Test (200-800)

MTY = race (1 = black, Chicano, Indian; 0 = other)

NOV GRADE = grade in November of first semester of Economics 10

OLS = ordinary least squares

PED = innovation (1 if participating, 0 otherwise)

[PED] = educational inputs other than instructor

PERINT = personal interests (1 = very important; 4 = not at all)

P/F = pass-fail (1 = yes; 0 = no)

POLITINT = political interests (1 = very important; 4 = not at all)

PREMED = premedical student

PROF/ENTR = entrepreneurial interests (1 = very important; 4 = not at all)

#PS = quality and quantity of problem sets and handouts; (number of problem sets varies)

RATES1 = secondary school teachers' ratings of a student's scholastic promise

RATES2 = Harvard admissions ratings of a student's scholastic promise

%READING = percent of reading assignments done

RELTIME = time spent on Economics 10 relative to other courses

RPD = not in class to receive questionnaire (0 = no questionnaire; 1 = questionnaire)

SATCOL (same

as COLSAT) = satisfaction with college life

(1 = not at all; 4 = very much)

SAT = intellectual satisfaction

SECON = economics concentrator (1 = economics concentrator; 0 = other)

SEE = standard error of estimate

SEX = 1 if female; 0 if male

SHUMAN = plan to concentrate in humanities

SMART = 1 if MSAT at least 750 and SUMS = 3, 4

SPI = self-paced section (1 = yes; 0 = no)

SUMS = preparation in mathematics in high school (1-4)

T = student time

TEACH = instructor characteristics

[TEACH] = vector of instructor characteristics

TEACHCOMP = intellectual competence of teacher (1 = high; 7 = low)

TEACHEX

(same as TEX) = teaching experience (number of years)

TEACHGRADES = grades of instructor in college and graduate school (0-15)

TEACHNICE = human qualities of instructor (1 = high; 7 = low)

TEX = teaching experience (number of years)

TFSPEC = intellectual experience

[Ts] = vector of student tastes

TUCE = Test of Understanding in College Economics

U^i = utility received by i th student

VSAT = score on verbal portion of Scholastic Aptitude Test (200-800)

FOOTNOTES

1. With self-paced instruction (SPI) the student is given a list of objectives for each unit of the course. Whenever a student believes he has mastered a unit he comes in and takes an exam on that unit. The exam is graded pass/fail. If he fails, there is no penalty; however, he must return to take a similar exam on the same unit. A student's grade depends only on the number of units he ultimately passes, not on the number of attempts. The purpose of SPI is to provide the student with abundant, relatively nonthreatening feedback.
2. For examples of this approach see any issue of the *Journal of Economic Education* or any empirical paper on economic education published in the May *Proceedings* issue of the *American Economic Review*.
3. Another possible source of bias lies in the errors-in-variables problems created by the imperfection of the observed effort term or ability term. This problem is discussed at greater length in Section II.
4. This is merely a translation from conventional production theory where a behavior function is introduced to take account of the maximizing process of the firm; i.e., of the possibility that disturbances are "transmitted." See Y.

Mundlack and I. Hoch, "Consequences of Alternative Specifications in Estimation of Cobb-Douglas Production Functions," *Econometrica*, October 1966.

5. One can, of course, treat institutional features in the manner of the Chicago school, i.e., as parameters in the maximizing process. It seems more useful, however, to be less dogmatic about the pervasiveness of the maximizing process and allow institutional features to enter directly.
6. This formulation assumes that first-semester grades within the course wash out the effects of previous grades in other courses.
7. There is not a priori reason to assume that a linear specification of the effort or enjoyment equations is appropriate either. Examination of the residuals with a variety of functional forms, however, suggested that a linear specification was appropriate.
8. There are other ways of relaxing this assumption; for example by the introduction of interaction terms or (as is done in Section IV of this paper) piecewise estimation.
9. Within social psychology there is a literature on the satisfaction of the college student which could be used to refine this formulation.
10. J. Kmenta, "On Estimation of the CES Production Function," *International Economic Review*, June 1967.
11. It is encouraging that the constellation of tastes thus defined corresponds closely to those used in several major studies of college students' interests and priorities.
12. In the Harvard system, students who wish to concentrate in economics are strongly urged to take the course in the freshman year; those who suspect they will not like the "economic way of thinking" typically put it off. Thus year of enrollment is a good guide to interest.
13. The imperfections of these measures of "analytic ability" are potentially the source of bias in the effort coefficients if this largely "unobservable" component of ability is correlated with effort.
14. TEACHNICE was not included because it was clearly arrayed along another dimension.
15. For a discussion of this and other approaches to the errors-in-variables problem, see Zvi Griliches, "Errors in Variables and Other Unobservables," *Econometrica*, November 1974.
16. CMI is consistently insignificant and is omitted in other achievement equations.
17. Some evidence for this hypothesis emerges when the data set is stratified by ability. For the top quarter, effort and enjoyment are positively (although insignificantly) correlated; the correlation is negative and significant for the bottom quarter.
18. The inappropriateness of this assumption is, for example, cited by Bowles as a major objection to the use of this function for analyzing data from primary and secondary schools (see Samuel Bowles, "Toward an Educational Production Function," in *Education, Income and Human Capital*, edited by W. Lee Hansen [New York: National Bureau of Economic Research, 1970]).
19. This result holds even when the dependent variable is not the score on all questions but the score on analytic questions only.
20. For a survey of the issues and evidence; see "Male-Female Differences in Economic Education: A Survey," John Siegfried, *Journal of Economic Education*, Spring 1976.
21. The question of why there might be a systematic difference in analytic power would then, of course, become the central question.
22. This may be one explanation for the insignificance of CMI.

THE EFFICIENCY OF EDUCATION IN ECONOMICS

THE EFFICIENCY OF PROGRAMMED LEARNING IN TEACHING ECONOMICS: THE RESULTS OF A NATIONWIDE EXPERIMENT*

By RICHARD E. ATTIEYEH, *University of California, San Diego*, G. L. BACH, *Stanford University*
and KEITH G. LUMSDEN, *Stanford University*

Summary

Recently considerable interest has developed in the use of programmed learning in the field of economics. While previous experiments¹ have been on a small scale, the results suggest that programmed texts can be very effective teaching materials. This paper reports the results of a larger, nationwide experiment, involving forty-eight schools and 4,121 students, designed to evaluate the efficiency of programmed materials in teaching the core micro- and macroeconomics sections of the typical elementary economics course. The major results of the study are:

1. On average, by spending twelve hours studying a programmed learning text students learned practically as much micro- or macroeconomics as did students in seven weeks of a conventionally taught elementary course.

2. On the basis of the test question breakdowns, students who used only programmed learning materials, as compared to conventionally taught students, performed better on "applications" of theory than on simple "concept recognition."

3. Students had a generally positive attitude toward programmed learning.

* The authors are grateful to the Joint Council on Economic Education for financial support, to Prentice-Hall, Inc., and McGraw-Hill, Inc., for providing programmed materials at cost, and to participating colleges and universities for their unstinted cooperation.

¹ See K. G. Lumsden, "The Effectiveness of Programmed Learning in Elementary Economics," *A.E.R.*, May, 1967, and "Technological Change, Efficiency and Programming in Economics," in *New Developments in the Teaching of Economics*, ed. K. G. Lumsden (Prentice-Hall, Inc., 1967).

These findings suggest that the basic concepts and tools of micro- or macroeconomics can be self-taught in about two weeks' time with programmed learning materials, thereby freeing a much larger portion of the total course to help students develop skills in the application of the basic theory to "real world" problems.

Experimental Design

The principal objective of this experiment was to compare the performance of students using programmed learning, either by itself or as a supplement, with that of students taking a conventionally taught elementary course. Student performance was measured by test scores on an independently devised, nationally normed multiple choice examination taken when the micro- or macroeconomics topics were completed. To separate the effects of the programmed materials from other variables, information was also obtained on the educational level, sex, and scholastic aptitude of students, the type, size, and quality of schools attended, and the textbook, class size, and experience of teacher for conventional sections.

Each participating school established three test groups. Students in Group I were given copies of one of two programmed texts and were told to study this text exclusively (avoiding other texts and not coming to class) for a period of time determined independently by each school.²

² The programmed texts used were R. C. Bingham, *Economic Concepts: A Programmed Approach* (McGraw-Hill, Inc., 1966) which contains both micro and macro sections; or either K. G. Lumsden, R. E. Attiyeh, and G. L. Bach, *Microeconomics: A Pro-*

AMERICAN ECONOMIC ASSOCIATION

On average, students in this test group had three weeks to read the programmed book, of which they used, according to their responses on a questionnaire, only twelve hours. At the end of this period they were tested, and then rejoined or became a conventional class.

In both Groups II and III, students were given conventional reading assignments and attended class lecture and discussion sessions. Students in Group II, however, were also required to read a programmed book, generally at a time and pace of their own choosing, while students in Group III were asked not to use the programmed book. In each school these two groups were tested at the same time, usually several weeks later than Group I. Each school had as much time as it wanted to cover the basic micro- or macro-economic analysis. On average, students in these two groups had seven weeks to prepare for the examination and students in Group II spent eight hours reading the programmed book.

The tests used were preliminary forms (two micro and two macro) of the *Test of Understanding in College Economics* (TUCE) prepared at the suggestion of the A.E.A. Committee on Economic Education by a special committee sponsored by the Joint Council on Economic Education.³ To avoid possible contamination of

the results for Groups II and III from the earlier testing of Group I, the test form for these two groups was different from that given to Group I.

The sample was drawn to provide adequate coverage for five main types of colleges and universities—"high prestige" schools, liberal arts colleges, large universities, state colleges (often formerly called teachers colleges), and junior colleges.⁴ Of the 4,121 students on whom we were able to acquire a complete file of data, 84 percent were male. The average student was a sophomore and had the equivalent of an SAT composite score of 1108. The average school had 10,620 students and an entering class with the equivalent of an SAT average composite score of 1122. The typical student in Groups II and III was in a class of seventy-seven, and had an instructor who had been teaching for 6.8 years. Sixty percent of the students in these two groups used one of three widely known textbooks, referred to as textbooks A, B, and C. The remaining 40 percent used one of ten different books, none of which was treated separately because of the small sample of students using each.

on the tests. Other members were G. L. Bach, William G. Bowen, Paul L. Dressel (Executive Director), R. A. Gordon, Bernard F. Haley, Paul A. Samuelson, John M. Stalnaker (Consultant), and George J. Stigler.

⁴Participating colleges and universities, some of which participated in both micro and macro portions of the study, were the following: "High-prestige" schools: Amherst, Dartmouth, Harvard, Haverford, Oberlin, Swarthmore, and Yale; liberal arts colleges: Allegheny, Bowdoin, Denison, Lafayette, Union, and Wesleyan; large universities: Case-Western Reserve, City College of New York, Illinois, Indiana, Iowa State, Johns Hopkins, Minnesota, North Carolina, Oklahoma, Oklahoma State, Pennsylvania, Rochester, University of California, San Diego, Utah, and West Virginia; state (teachers) colleges: Arizona Western, California State, Fullerton, Georgia State, New Mexico State and San Diego State; junior colleges: Allegheny, Bakersfield, Chabot, Cochise, Everett, Hill, Phoenix, Riverside City, Robert Morris and Triton; norm group: Butler, Geneva, Montana, State University of New York at Albany and Virginia Polytechnic Institute.

grammed Book (Prentice-Hall, Inc., 1966) or R. E. Attiye, K. G. Lumsden, and G. L. Bach, *Macroeconomics: A Programmed Book* (Prentice-Hall, Inc., 1967). Approximately two-fifths of the students in Group I used the Bingham text; one-third of the students in Group II used the Bingham text. Others used one of the Attiye, Lumsden, and Bach texts. All programmed learning and test materials were provided free to participating schools. Each school was allowed to choose the programmed text preferred, subject only to the constraint that each book be used by a typical cross-section of schools.

³For permission to use these tests we are grateful to the Psychological Corporation (304 East 45th Street, New York, New York, 10017) and to Professor Rendigs Fels, chairman of the TUCE committee, whose paper at this session provides more information

EFFICIENCY OF EDUCATION IN ECONOMICS

Results

The main findings of this study are reported in Table 2 which presents the statistics for a regression involving all of the variables discussed above. The dependent variable—the number of correct answers on the TUCE—had a mean of 17.90 and a maximum value of 30.⁶

All of the variables representing student characteristics were statistically significant⁷ and quantitatively important. Each year of educational attainment added .39 points to the predicted grade, while being a girl cost .75 points. Not unexpectedly, student ability was the most important single determinant of student performance. Each 100 points on the composite college entrance exam (SAT scale) was worth 1.03 correct answers on the TUCE exam.

The results for school characteristics were somewhat more surprising. First, the average freshman entrance exam score was highly significant and nearly half as important to an individual's performance as his own entrance exam score. Each 100 points of the average SAT score of the freshman class added .45 correct answers to the student's score, even after all other individual student qualities are taken into account. It is impossible to tell whether this gain was due to associating with bright students or attending a high quality school that attracts bright students.

Second, school size had a positive, statistically significant coefficient; each 10,000 in enrollment added 0.2 points to the predicted score. Since school size in our sample ranged from 500 to 45,000

students, going to the largest university rather than the smallest school added approximately .9 of a question to student performance. We have no satisfactory explanation. It may simply be that larger schools have better courses and/or better instructors as measured by the TUCE. Or, perhaps, students who attend larger schools put their talent to better use.

Finally, of the school type variables, prestige schools and state colleges had large and statistically significant coefficients. Other things the same, compared to "other" schools, attendance at a "high prestige" school cost .78 points while being at a state college added 1.30 points. Whether these coefficients reflect differences in teaching effort or some other pertinent factor is difficult to say. In evaluating how a student would perform at schools of different types, however, it should be borne in mind that the quality (as measured by the average freshman SAT score) and size of these schools must also be taken into account. Leaving these factors free to vary, the influence on test score of attending a school of each type, as compared to "other" schools, is as follows: prestige schools, +.57; large state universities, +.71; liberal arts colleges, +1.08; state colleges, +.83; and junior colleges, -.90. These numbers still suggest that a student would perform better on the TUCE if he had gone to a less prestigious liberal arts college, a state college or a large university than if he had gone to a prestige college or university.⁷

Variables 11-13 show the difference in

⁶ Actually, the TUCE consists of 33 questions. On the macro forms, however, we deleted from each form 6 questions dealing with microeconomics, leaving 27 questions. Two tests, therefore, had 33 questions each and two had 27 each. Thus, for all four forms, the unweighted average maximum score was 30.

⁷ A confidence level of 95 percent will be assumed throughout the following discussion.

⁷ A similar classification was used in the study by G. L. Bach and Phillip Saunders entitled "Lasting Effects of Economics Courses at Different Types of Institutions," *A.E.R.*, June, 1966. In this study the significant finding was that the largest lasting effects for elementary economics courses were traceable to liberal arts colleges and the least lasting effects to large universities. Those results measure performance eight years after high school economics teachers took the elementary course.

AMERICAN ECONOMIC ASSOCIATION

TABLE 1

REGRESSION RESULTS

Dependent variable: test score (mean = 17.90)

Coefficient of determination = .45

Standard error of estimate = 3.51

Independent Variable	Mean	Regression Coefficient	t^{**} Statistic
Intercept		.28	
1. Educational level (year)	1.94	.39	5.34
2. Sex	1.16	-.75	-5.01
3. Entrance exam (SAT, hundreds)	11.08	1.03	28.37
4. Average school entrance exam (SAT, hundreds)	11.22	.45	5.95
5. School size (thousands)	10.62	.02	2.33
6. "Prestige" schools	.16	-.78	-2.40
7. Large state universities	.40	-.03	-.11
8. Liberal arts colleges	.13	.28	.92
9. State colleges	.08	1.30	4.54
10. Junior colleges	.15	.00	.00
11. Microeconomics test A	.24	.55	3.13
12. Microeconomics test B	.27	2.47	14.45
13. Macroeconomics test A	.27	-.31	-1.94
14. Conventional text A	.27	.86	4.68
15. Conventional text B	.09	.42	1.73
16. Conventional text C	.10	.75	2.89
17. Years teaching experience	4.67	.01	.26
18. (Years of teaching experience) ^a	68.96	-.00	-.49
19. 1+class size	.02	.30	.06
20. P. L. book A+conventional	.26	.55	3.23
21. P. L. book B+conventional	.13	-.37	-1.72
22. P. L. book A	.19	-.02	-.07
23. P. L. book B	.12	-.46	-2.02

performance on the four test forms used. The standard of comparison is macroeconomics test B. Thus macroeconomics test A is .31 of a question more difficult. Since the microeconomics tests, however, each had 6 questions more and since students on average answered more than 50 percent correct (i.e., 17.9 out of 30), we would expect, if the tests were homogeneous, that the coefficient on each of the micro tests would be greater than 3.0 (i.e., the average student would be correct on just over 3 of the extra 6 questions). Thus microeconomics test A is considerably more difficult than macro B which was the easiest of all the tests. It should be re-

called that preliminary forms of the tests were used.

A surprising result is that teaching experience and class size have no observed effect on student performance.^a What textbook is used, however, does appear to be a matter of some importance. Both texts A and C had positive significant coefficients. Other things equal, students who used these texts scored .86 or .75 points higher than students who used the average book in the "all other" category. There is no

^a It should be noted that the result holds even though we tested for nonlinear relationships. We recognize, however, that experience is not necessarily a good proxy for quality of teaching.

EFFICIENCY OF EDUCATION IN ECONOMICS

TABLE 2
PROGRAMMED LEARNING COMPARED TO CONVENTIONAL INSTRUCTION

Alternative to Conventional Instruction	Difference in Predicted Score (P.L.-average conventional)	"t" Statistic*
P.L., only, book A.....	-0.54	-1.78
P.L., book A and conventional	0.55	3.23
P.L., only, book B.....	-1.02	-4.18
P.L., book B and conventional	-0.37	-1.72

* Calculated from Table 2 and the underlying variance-covariance matrix of the regression coefficients.

significant difference, however, between either books A or C and book B.

In evaluating the effectiveness of programmed learning taken by itself, the results are somewhat difficult to read from Table 1. The coefficients shown in Table 1, $-.02$ for P.L. book A and $-.46$ for P.L. book B, reflect the difference between programmed learning and a conventional course, assuming that variables 14-19 have a value of zero. In the case of textbooks, for example, this means that students in Group III are assumed to be using a text other than books A, B, or C. To make a comparison with the typical conventional course, it is necessary to give appropriate weight to the coefficients for variables 14-19 in Table 1. This comparison is shown in Table 2. What these figures indicate is that, other things the same, students using programmed learning did half a question worse, in the case of P.L. book A, and one question worse in the case of P.L. book B, than did students in the average conventional course. In the case of P.L. book A, however, this difference is not statistically significant.

As for the effectiveness of programmed learning when it is used as a supplement, P.L. book A added significantly, about half a question, to student performance, whereas P.L. book B did not. These results are shown in both Tables 1 and 2.

The questions on the TUCE were clas-

sified into the following categories: recognition and understanding; simple application; and complex application. Questions in the first category required knowledge of relatively more historical and institutional material, while questions in the latter two categories usually stated all the necessary factual material and tested the ability to use economic analysis. Separate regressions were run to explain student performance on each type of question. The statistics for these regressions are shown in Table 3 and their implications for the comparison between programmed learning and conventional instruction are shown in Table 4. The most striking result is that, for P.L. book A, most of the overall difference between student performance in Groups I and III was concentrated in the recognition and understanding category. Since the programmed learning texts were designed primarily to teach basic tools and not all basic introductory and institutional material covered by the typical course or the TUCE, the differences in this category are not surprising. For those questions requiring an ability to apply economic theory, however, students who used P.L. book A only for 12 hours, on average, were an even match for conventionally taught students.

In addition to the objective information already discussed we solicited student reactions to programmed learning. The re-

AMERICAN ECONOMIC ASSOCIATION

TABLE 3
REGRESSION RESULTS BY TYPE OF QUESTION

Independent Variables	Recognition and Understanding		Simple Application		Complex Application	
	Regression Coefficient	"t" Statistic	Regression Coefficient	"t" Statistic	Regression Coefficient	"t" Statistic
Intercept.....	1.35		-.05		.66	
1. Educational level (year).....	.12	3.91	.12	3.67	.19	5.22
2. Sex.....	-.34	-5.44	-.21	-3.02	-.27	-3.50
3. Entrance exam (SAT, hundreds).....	.28	18.20	.37	22.12	.38	20.15
4. School entr. exam. (SAT, hundreds).....	.12	6.08	.08	3.80	.11	4.52
5. School size (thousands).....	.01	4.63	.01	2.29	.00	1.34
6. Microeconomics test A.....	.04	.62	1.33	17.15	-.75	-8.51
7. Microeconomics test B.....	-.07	-.96	2.57	34.41	-.07	-.85
8. Macroeconomics test A.....	.83	12.37	-.19	-2.64	-.96	-11.46
9. Conventional text A.....	.26	3.47	.14	1.71	.27	2.92
10. Conventional text B.....	.28	2.88	.10	.97	.26	2.18
11. Conventional text C.....	.31	3.12	-.05	-.43	.07	.59
12. P.L. book A+conventional.....	.19	2.65	.14	1.76	.21	2.41
13. P.L. book B+conventional.....	-.09	-1.03	-.14	-1.54	-.10	-.94
14. P.L. book A.....	-.24	-2.62	-.05	-.49	.10	.92
15. P.L. book B.....	-.15	-1.76	-.24	-2.53	-.27	-2.46
	$R^2 = .31$ $S = 1.47$		$R^2 = .45$ $S = 1.61$		$R^2 = .31$ $S = 1.83$	

sponses to the questions concerning effectiveness and interest are summarized in Table 5. Several aspects of these distributions stand out. First, the mean response was favorable, but there was a wide dispersion of opinions. Second, students con-

sidered programmed learning to be somewhat more effective than interesting. On average, the students gave the books grades of "good-minus" for effectiveness and "average-plus" for interest. Finally, although the objective evidence revealed a

TABLE 4
PROGRAMMED LEARNING COMPARED TO CONVENTIONAL INSTRUCTION, BY TYPE OF QUESTION
Difference in predicted score* (P.L.—average conventional)

Alternative to Conventional Instruction	Recognition and Understanding "t"		Statistic Application "t"		Complex Application "t"	
	Difference	Statistic	Difference	Statistic	Difference	Statistic
P.L. only, book A.....	-.42	-4.29	-.11	-1.03	.02	.17
P.L. A and conventional....	.19	2.65	.14	1.76	.21	2.41
P.L. only, book B.....	-.34	-3.58	-.32	-3.08	-.36	-3.05
P.L. B and conventional....	-.09	-1.03	-.14	-1.54	-.10	-1.13

* Calculated from Table 5. The "t" statistics are calculated using information from the variance-covariance matrix of the regression coefficients for the regression reported in Table 3.

EFFICIENCY OF EDUCATION IN ECONOMICS

TABLE 5
STUDENT ATTITUDES TOWARD PROGRAMMED LEARNING*

As a way of learning economic theory what do you think of the programmed text?

	Group I	Group II
Very good.....	20%	18%
Good.....	42	43
Average.....	23	25
Poor.....	12	10
Very poor.....	3	4

How interesting was the programmed text?

	Group I	Group II
Very interesting.....	6%	3%
Interesting.....	41	32
Average.....	34	42
Uninteresting.....	12	15
Boring.....	6	7

* Figures may not add to 100 because of rounding or failure to receive answers.

significant difference between the two books, this was not reflected in the student's opinions about them.

Conclusion

We feel that these results have important implications for the organization and teaching of the introductory course. Within the profession many believe that

the introductory course should prepare a student to think intelligently about major economic problems in modern society and that this goal can best be accomplished by teaching a few basic principles and applying them to a number of important problems. We are in agreement with this view. This study has shown that by using programmed learning materials the basic micro- and macroeconomic theory can be taught in a relatively short period of time. Therefore, more time can be devoted to teaching students how to apply the theory to social problems, both by going more deeply into the more important problems and by actually covering those topics scheduled for the end of the term that often fall victim to the school calendar. The use of these materials can have other advantages: First, the student can gain a good overview of the entire course at the very beginning which helps him to put topics covered in the remainder of the course in meaningful perspective. Second, because a course taught in this manner emphasizes the usefulness of economic theory in a problem solving context, it promises a positive impact on the most important single factor in the learning process; namely, student attitude toward the subject.

The Learning and Cost Effectiveness of AVT Supplemented Instruction: Specification of Learning Models

William E. Becker and Michael K. Salemi

The self-paced audiovisual tutorial (AVT) approach of instruction has been acclaimed by many educators as an efficient method of increasing student learning.¹ The rationale given for expecting increased learning includes the belief that:

1. Instructors will be able to extricate themselves from transmitting simple information and theory, thereby allowing themselves to spend time on higher levels of application; and
2. Students will be able to allocate their time to those areas which they find difficult and away from those of less difficulty.²

The research described here is designed to examine whether the integration of an AVT program in the Community College economic principles course does affect the quantity, cost and efficiency of student learning.

This research endeavors to answer four questions. First, can a learning model be specified, on the basis of formal theoretical and statistical grounds, within which learning can be examined in control-experimental groups? Unlike earlier studies, this study develops a theoretical and statistical model for equations estimated.³ Second, is there any difference in the quantity of learning as measured by the Test of Understanding in College Economics (TUCE) between control groups and experimental groups using the AVT approach? This question is typical of those raised in control-experimental testing.⁴ Third, what influence does student classroom and study time have on learning? In this study explicit account is taken of the fact that student time is an input to student learning.⁵ Fourth, is learning produced in the control sections? Unlike most other economic education studies, this study utilizes human capital techniques to assign value to time.

The sample data for this study were collected during the 1972-74 academic year at six urban and rural community colleges—three located in Minnesota and three in Missouri. At each college, the same instructor taught both a control section using his regular teaching plan and an experimental section in which student use of the David A. Martin AVT package, *Introductory Economic Theory*, was completely substituted for the instructor's presentation of several theoretical topics.

William E. Becker is Associate Professor and Director of Economic Education, University of Minnesota, Minneapolis, and Michael K. Salemi is Assistant Professor of Economics, University of North Carolina. They accept jointly responsibility for all aspects of the research. Darrell Lewis and Bill Walstad provided constructive criticism of an earlier version of this paper. DeVon Yoho deserves special recognition for his efforts in arranging for Missouri Community Colleges which participated in this study. This research was supported by a grant from the Exxon Foundation and the Joint Council on Economic Education.

In the first phase of the study it is hypothesized that community college economic principles course learning (measured as a student's post-TUCE score minus his pre-TUCE score) is linearly related to aptitude (as measured by the pre-TUCE score), a student time input (as reported in weekly questionnaires) and a set of dummy variables characterizing the students and their learning situation.⁶ The hypothesis that there is *no* difference in learning between control and experimental groups is accepted. Also it is shown that student study time has little effect on learning and that learning and pre-TUCE scores are negatively correlated.

Other researchers have been confronted with a negative coefficient for pre-TUCE in similar "value added" learning models and have related this to "nonlinearity" and the TUCE "ceiling effect."⁷ These researchers usually specify alternative *ad hoc* specifications.⁸ In the second phase of this study, it is shown that a formal modeling of a ceiling effect with pre-TUCE as a positive regressor in a "value-added model" can be accomplished, that such a modeling involves simultaneous equations in an unobserved component, and that an implication of this approach is that ordinary least square regression (OLS) estimates are biased and inconsistent. An instrumental variable procedure is undertaken to estimate a nonlinear learning model and to retest the hypothesis that there was no difference in learning (change in TUCE) between control and experimental groups.

In addition to examining the quantity of learning produced in control and experimental sections, the study attempts to determine whether the costs of learning are different in the control and experimental groups. An appropriate cost concept is developed and the cost of learning is examined for control and experimental sections; the null hypothesis that there is *no* difference in the average cost of learning per TUCE point between the two groups is accepted at the 0.05 Type I error level.

Experimental Design

This section briefly describes both the sample and course design used. For a complete description of operational procedures used in data collection and a detailed course description, the reader is advised to obtain our 1973-1976 yearly program reports to the Joint Council on Economic Education.⁹

Sample

This study involves 330 students from six two-year community colleges, three in Minnesota and three in Missouri. Although these schools differed in terms of geographic area, they were quite similar in terms of philosophy, objectives, resources, type of student body and type of introductory economics course.

The Minnesota courses were quarter courses while the Missouri courses were semester courses. Both the quarter and semester courses, however, covered the same amount of materials in roughly the same amount of time. The instructors of each school's introductory economics course were quite similar in terms of academic credentials, teaching experience and "willingness" to become involved in a study of the AVT method of instruction. Each of the six instructors volunteered to teach both a control and an experimental class at his respective school.¹⁰

All students were required to take a pre- and postcourse test of their understanding of college economics (TUCE). They completed a pre- and postcourse questionnaire identifying personal characteristics and attitudes. On the basis of the prequestionnaire data, Walstad [35] found that students in the control and experimental classes of each school were similar in terms of key characteristic measures (age, sex, aptitude, educational background, outside employment, etc.). Students as well as instructors were also required to complete weekly question-

naires reporting on time usage during the week.

Each instructor carried out the administration and collection procedures at his institution. Since the questionnaire and test-scoring was done at the University of Minnesota and because special handling and mailing procedures were used, individual instructors did not have access to raw data.

Experimental Course

Before starting the program in the winter of 1974, the six instructors received background information about the program, attended a special workshop to learn about the AVT package to be used and became acquainted with the procedures of this study. Using the David Martin, *Introductory Economic Theory*, audiovisual-tutorial package, the six instructors, together with their school's media experts, set up study areas at each of the schools. Each school's study area contained several tape playback machines, slide projectors, viewing screens and study carrels. Several copies of the Martin material (slides, carrousels and audiotapes) were kept in the study area with each experimental student receiving his or her own copy of the Martin workbook. The study areas were centrally located on campus and had a wide range of hours during which they were open.

Each experimental course was designed to be nine classroom sessions shorter than the control course and each instructor was free to schedule the omitted sessions. Content covered in units two, three, four and five of the Martin package was *not* explicitly covered in the experimental classrooms, so that experimental group students learned the material in these units at their own initiative and on their own time with the aid of the Martin AVT package. Each instructor could make use of the omitted content in classroom discussions of current issues but was asked not to give classroom assistance or explicit attention to teaching the content and concepts covered in the relevant Martin units! Students were permitted to see the instructors for individual help outside of the classroom. Except as provided above each instructor's control and experimental classes were taught in a similar way.

Experimental Results with Linear Learning Models

In general, this section and the following one focus on a relatively simple model of learning:

$$(1) \quad L = f[A, T, S, u]$$

where L = Learning

A = Aptitude

T = Time input

S = Situation

u = Random error

The model says that learning is to be explained as dependent, except for an error component, on a student's aptitude, the time the student spends in class and in study, and the situation or environment within which he learns. The function f may be thought of as a production function with A a measure of human capital, T a measure of labor time, and S characterizing the "physical learning plant." Variables such as age and sex are not included explicitly inasmuch as their contribution, if any, should be reflected in A .¹¹ Within the context of specific versions of this model the question is asked: "Did participation in the experimental sections (those with the Martin AVT component) make a difference in learning?"

Estimation (Time Suppressed)

The theoretical learning function given in (1) can take numerous algebraic forms; a linear model with time temporarily suppressed is:

$$(2) \text{ D TUCE} = \sum_{i=1}^6 \sum_{j=0}^1 \alpha_{ij} \text{ GROUP}(i, j) + \sum_{i=1}^6 \sum_{j=0}^1 \beta_{ij} \text{ GROUP}(i, j) \cdot \text{Apt} + u$$

where D TUCE = the difference between post-TUCE and pre-TUCE scores

$$\text{GROUP}(i, j) = \begin{cases} = 1 & \text{if a student attended school } i \text{ and participated in section } j. \\ & (j = 0 \text{ for control, } j = 1 \text{ for experimental}). \\ = 0 & \text{otherwise} \end{cases}$$

Apt = pre-TUCE

Learning is measured by D TUCE while pre-TUCE is used as a proxy for the students' unobserved aptitude to learn economics. A student's learning situation is described by the school which he attended and the section in which he was placed. Aptitude enters the equation in a way which would permit the discovery of interaction with the situational variables if any exists. Table 1 summarizes the OLS regression results:

The null hypothesis that participation in the experimental AVT sections had no effect on learning is tested against the alternative hypothesis used in estimating (2). The appropriate constrained regression to estimate is:

$$(3) \text{ D TUCE} = \sum_{i=1}^6 \alpha_i \text{ School}(i) + \sum_{i=1}^6 \beta_i \text{ School}(i) \cdot \text{Apt}$$

where School(*i*) = 1 if the student attended School(*i*)
0 otherwise

Apt = pre-TUCE

Interaction is once again permitted between the aptitude proxy variable pre-TUCE and the situational variable School(*i*) (see Table 1).

The *F*-statistic for the test of the null against the alternative hypothesis was estimated to be 1.26 which is not significant at the .05 level. In fact, $P\{\bar{X} \sim F(12, 306) \geq 1.26\} = .24$. The hypothesis that participation in the experimental section had no differential effect on learning cannot be rejected. It is interesting to observe, however, that learning is significantly different across schools. In addition, a striking feature of the results is the ever-present negative coefficient on pre-TUCE which casts doubt on its role as a measure of aptitude in the system as written in (2).

Estimation (Time as an Input)

Since it is taken to be the labor input of a learning production function (1), the time a student devotes to the economics course should have an effect on learning. A linear learning model, given the hypothesis that time is an input to learning, is:

$$(4) \text{ D TUCE} = \sum_{i=1}^6 \sum_{j=0}^1 \alpha_{ij} \text{ GROUP}(i, j) + \beta_{ij} \text{ GROUP}(i, j) \cdot \text{Apt} \\ + \gamma_{ij} \text{ GROUP}(i, j) \cdot \text{TIME} + u$$

GROUP(*i, j*) and Apt are defined as before and TIME is defined as the course average of weekly responses to the questionnaire item "How much total time (classroom time, reading time, study time, etc.) did you devote to this economics course this past week?"

In this case, the student's time input in the course is allowed to explain learning and interaction between time input and the situational variable is permitted. The appropriate constrained equation for a test of the null hypothesis that participation in the experimental

Table 1
Dependent Variable = D TUCE

Variable	Coefficient	Standard Error
Unconstrained Regression (2)		
GP (0, 1)	5.71	2.83
GP (1, 1)	7.18	3.06
GP (0, 2)	7.68	2.74
GP (1, 2)	9.62	3.11
GP (0, 3)	7.87	2.49
GP (1, 3)	5.14	2.54
GP (0, 4)	11.05	2.91
GP (1, 4)	7.50	3.15
GP (0, 5)	12.07	2.48
GP (1, 5)	10.52	2.63
GP (0, 6)	10.29	4.00
GP (1, 6)	14.55	3.16
PREGP (0, 1)	-.18	.21
PREGP (1, 1)	-.24	.22
PREGP (0, 2)	-.18	.20
PREGP (1, 2)	-.42	.26
PREGP (0, 3)	-.43	.19
PREGP (1, 3)	-.29	.18
PREGP (0, 4)	-.77	.24
PREGP (1, 4)	-.25	.27
PREGP (0, 5)	-.81	.21
PREGP (1, 5)	-.60	.22
PREGP (0, 6)	.11	.34
PREGP (1, 6)	-.49	.27

$R^2 = .345$

$F(24,306) = 19.85$

Residual sum of squares = 5296

$N = 330$

Constrained Regression (3)		
Sch 1	6.26	2.07
Sch 2	7.91	1.98
Sch 3	6.65	1.78
Sch 4	9.80	2.14
Sch 5	11.32	1.81
Sch 6	12.94	2.49
PRESCH 1	-.20	.15
PRESCH 2	-.23	.16
PRESCH 3	-.38	.13
PRESCH 4	-.55	.18
PRESCH 5	-.71	.15
PRESCH 6	-.25	.21

$R^2 = .313$

$F(12,318) = 38.07$

Residual sum of squares = 5558

$N = 330$

section had no differential effect on learning is:

$$(5) \text{ D TUCE} = \sum_{i=1}^6 \alpha_i \text{ School } (i) + \beta_i \text{ School } (i) \cdot \text{Apt} + \gamma_i \text{ School } i \cdot \text{Time} + u$$

The OLS regression results are summarized in Table 2.

Table 2
Dependent Variable = D TUCE

Variable	Coefficient	Standard Error
Unconstrained Regression (4)		
GP (0, 1)	8.65	3.67
GP (1, 1)	3.07	3.55
GP (0, 2)	5.61	3.14
GP (1, 2)	12.88	4.25
GP (0, 3)	6.44	3.65
GP (1, 3)	5.02	2.71
GP (0, 4)	8.42	5.11
GP (1, 4)	2.45	4.78
GP (0, 5)	16.48	3.76
GP (1, 5)	7.77	3.97
GP (0, 6)	12.44	8.10
GP (1, 6)	8.84	4.86
PREGP (0, 1)	-.20	.21
PREGP (1, 1)	-.36	.22
PREGP (0, 2)	-.19	.20
PREGP (1, 2)	-.49	.27
PREGP (0, 3)	-.43	.19
PREGP (1, 3)	-.29	.18
PREGP (0, 4)	-.73	.25
PREGP (1, 4)	-.31	.27
PREGP (0, 5)	-.90	.21
PREGP (1, 5)	-.57	.22
PREGP (0, 6)	.19	.41
PREGP (1, 6)	-.49	.26
TIMEGP (0, 1)	-.42	.34
TIMEGP (1, 1)	1.00	.45
TIMEGP (0, 2)	.34	.26
TIMEGP (1, 2)	-.38	.34
TIMEGP (0, 3)	.24	.46
TIMEGP (1, 3)	.03	.22
TIMEGP (0, 4)	.42	.67
TIMEGP (1, 4)	1.14	.82
TIMEGP (0, 5)	-.65	.42
TIMEGP (1, 5)	.42	.46
TIMEGP (0, 6)	-.48	1.59
TIMEGP (1, 6)	1.07	.70

$R^2 = .38$

$F(36, 294) = 13.98$

Residual sum of squares = 4994

Table 2 (Continued)
Dependent Variable = D TUCE

Variable	Coefficient	Standard Error
Constrained Regression (5)		
Sch 1	5.98	2.58
Sch 2	7.24	2.46
Sch 3	6.26	2.06
Sch 4	5.64	3.50
Sch 5	11.92	2.73
Sch 6	6.72	3.86
PRESCH 1	-.20	.15
PRESCH 2	-.23	.16
PRESCH 3	-.38	.13
PRESCH 4	-.53	.18
PRESCH 5	-.71	.15
PRESCH 6	-.33	.21
TIMESCH 1	.05	.26
TIMESCH 2	.10	.21
TIMESCH 3	.08	.20
TIMESCH 4	.76	.51
TIMESCH 5	-.09	.31
TIMESCH 6	1.25	.59

$R^2 = .33$
 $F(18,312) = 25.87$
 Residual sum of squares = 5434

The F -statistic for the test of the null hypothesis is estimated to be 1.44 with $P \{ \bar{X} \sim F(18,294) \geq 1.44 \} = .11$. The null hypothesis that there is no differential effect cannot be rejected at the .05 level. In this case, the time input as measured in this study does not make a significant contribution to the explanation of learning. The F -statistic which tests the null hypothesis represented by equation (3) against the alternative (4) is estimated to be 1.39 with $P \{ \bar{X} \sim F(24,294) \geq 1.39 \} = .11$.

It is important to note that the above-given results do not mean that there was no statistically significant difference in learning between control and experimental sections at any of the six schools. Rather, the results say that taken together there was no difference in learning between control and experimental sections.

Nonlinear Learning Models and Correction for Simultaneous Equations Bias

In this section the implications of fitting a learning model like (1) are carefully studied. In particular, the learning equation is viewed as the reduced form of a simultaneous equations system in which an explicit relation is posited between aptitude and pre-TUCE. The fact that there is a maximum pre-TUCE score and a maximum to TUCE measurable learning is incorporated in the model.

Pre-TUCE as an Aptitude Proxy

The learning function in (1) related learning output (D TUCE) to the inputs aptitude, A , situation, S , labor time, T , and an error term, u .

$$(1) \quad D \text{ TUCE} = f(A, S, T, u)$$

In the previous section it was simply assumed that pre-TUCE was some type of proxy for

aptitude which is unobserved. At this point, the relationship between pre-TUCE and aptitude is assumed explicitly to be:

$$(6) \quad \text{pre-TUCE} = h (\text{Aptitude}, v)$$

where v is a random error.

Solving (6) for aptitude and substituting in (1) gives the reduced form

$$(7) \quad D \text{ TUCE} = k (\text{pre-TUCE}, S, T, w)$$

The reduced form error w is made up of the shocks u and v which implies, in general, that w will be correlated with pre-TUCE.

Assuming explicit linear equation forms, such as those used in the previous section, the two-equation system would be given by (8) and (9) with reduced form (10):

$$(8) \quad D \text{ TUCE} = a + b \text{ Aptitude} + c' \text{ Situation} + d \text{ Time} + u$$

$$(9) \quad \text{pre-TUCE} = \gamma + \theta \text{ Aptitude} + v$$

$$(10) \quad D \text{ TUCE} = a' + b' \text{ pre-TUCE} + c' \text{ Situation} + d \text{ Time} + w$$

where $a' = (a - \lambda b/\theta)$

$b' = b/\theta$, $\theta > 0$, $b > 0$

c' = a vector of coefficients conformable to the situation vector

$w = u - (b/\theta) v$

A necessary condition for OLS to give consistent estimates of the parameters in (10) is that regressors are uncorrelated with w . However,

$$E (\text{pre-TUCE} \cdot w) = E [(\lambda + \theta \text{ Aptitude} + b) (u - \frac{b}{\theta} v)] = - \frac{b}{\theta} E [v^2] < 0$$

Therefore, OLS estimates of b' are biased and inconsistent.¹² It is this source of simultaneous equation bias which may account for the highly significant negative pre-TUCE coefficient estimates reported in the previous section.¹³

An appropriate remedy for the problem of simultaneous equation bias is the use of an instrumental variable procedure such as two stage least squares (TSLS). The results of reestimating a basic linear learning model via TSLS are given in Table 3.

Table 3
Two Stage Least Squares Estimates of Linear Learning Model
Dependent Variable = D TUCE

Variable	Pre-TUCE	School 1	School 2	School 3	School 5	School 6	Conex Time
Coefficient	.22	-.17	2.45	-1.55	.40	6.69	-.56 .17
Standard error	.23	3.06	2.99	3.08	2.65	2.75	.63 .45
$R^2 = .11 \quad N = 219 \quad \text{Residual sum of squares} = 4421$							

It is worth noting that the TSLS pre-TUCE coefficient estimate is positive (although not significantly so) which is consistent with the assumption that pre-TUCE is a proxy for aptitude. Once again, however, there is no discernible difference between the learning in the control and experimental groups since the coefficient of the Control-Experimental Dummy Variable (Conex) is not significantly different from zero. The coefficient of student time is also estimated to be insignificantly different from zero.

The Gap-Closing Model

Researchers in models of economic learning have puzzled over the negative slope coefficient estimate for pre-TUCE in a regression of post-TUCE - pre-TUCE and other

variables. Some have suggested that the negative coefficient is evidence of a ceiling effect created by the fact that there is a maximum possible score on any testing instrument. But we have seen that the negative coefficient may be spuriously due to simultaneous equation bias. What follows is an attempt to model such an effect explicitly.¹⁴

It seems reasonable to assume that the first derivative of (6) with respect to aptitude is positive while the second is negative for pre-TUCE scores ranging over the set $[0, 33]$. An explicit equational form which has such characteristics is the logistic (11).

$$(11) \quad \text{pre-TUCE} = 33 \left[\frac{1}{1 + \frac{1}{A}} \right]^r e^v$$

where A = aptitude

v = random error

$e = 2.71828$

In (11) with $S > 0, A \rightarrow \infty$ gives pre-TUCE $\rightarrow 33$ at a decreasing rate over the relevant domain.

There is also a maximum to learning as defined in model (1); with a zero pre-TUCE one can at most gain 33 points with a perfect test score. Equation (12) reflects the fact that even with a very high aptitude and a very long effort a student will at best score well enough on the posttest to close the gap between a perfect score and his pretest score.

$$(12) \quad \text{D TUCE} = (33 - \text{pre-TUCE}) \left[\frac{1}{1 + \frac{1}{A}} \right]^r \left[\frac{1}{1 + \frac{1}{T}} \right]^z \left[\pi_{ij} (S_{ij})^{\rho_{ij}} \right] e^u$$

where $G_{ij} = 1$ if the student belongs to school i and participated in section j
0 otherwise

S_{ij} = the multiplicative effect on learning of belonging to group (i, j) ; $S_{ij} \geq 0$

With r greater than zero, as $A \rightarrow 0$, D TUCE $\rightarrow 0$ and as $A \rightarrow \infty$, D TUCE $\rightarrow$ TUCE Gap times a constant which permits the effect of aptitude on D TUCE to be different in different groups for a given time commitment. Similarly, with z greater than zero as $T \rightarrow 0$, D TUCE $\rightarrow 0$ and as $T \rightarrow \infty$, D TUCE $\rightarrow$ TUCE Gap times a constant which permits the effect of time on D TUCE to be different in different groups for given aptitude levels. As either A or T rise, D TUCE rises at a decreasing rate over the relevant domain.

Solving (11) for aptitude and substituting into (12) yields

$$(13) \quad \text{D TUCE} = (33 - \text{pre-TUCE}) \left[\frac{\text{pre-TUCE}}{33e^r} \right]^{\frac{r}{s}} \left[\frac{1}{1 + \frac{1}{T}} \right]^z \left[\pi_{ij} (S_{ij})^{\rho_{ij}} \right] e^u$$

Dividing both sides of (13) by $(33 - \text{pre-TUCE})$ and rearranging factors gives the familiar "Gap-Closing" measure as the dependent variable.¹⁵

$$(14) \quad \frac{\text{D TUCE}}{33 - \text{pre-TUCE}} = \left[\frac{\text{pre-TUCE}}{33} \right]^{\frac{r}{s}} \left[\frac{1}{1 + \frac{1}{T}} \right]^z \left[\pi_{ij} (S_{ij})^{\rho_{ij}} \right] e^{u - r/s}$$

In (14), it can be seen that the measure of learning, the fraction of the TUCE Gap closed, is positively related to pre-TUCE if $r/s > 0$.

The reduced form (14) can be estimated by performing TSLS on its base e logarithmic form:

$$(15) \quad \text{Ln} \frac{\text{D TUCE}}{33 - \text{pre-TUCE}} = \frac{r}{s} \left[\text{Ln} \frac{\text{pre-TUCE}}{33} \right] + z \left[\text{Ln} T - \text{Ln} (T + 1) \right] + \sum_i \text{Ln}(S_{ij}) f_{ij} + (u - rv/s)$$

The test that there was no difference in learning between control and experimental groups is the test that $S_{0i} = S_{1i}$, $i = 1, \dots, 6$.

The Gap-Closing model is misspecified in at least one important way, however. As written, (15) implies that random shocks scale up or down the effects of the explanatory variables and, thus, the model predicts that D TUCE will be signed identically for all students in group (i, j) . If $S_{0j} > 0$, the model predicts $\text{D TUCE} \geq 0$. But this is a serious inadequacy since a student may guess the answer to questions he or she does not know. It is possible that a student who learned little and guessed a lot might score higher on the pre-TUCE than on the post-TUCE. It is also possible but was not the case that such a guessing phenomenon may be more likely at lower aptitude levels.¹⁶

The scheme used to estimate (15) is simply to drop from the sample those cases for which $\text{D TUCE} \leq 0$. As suggested above, this is done on the grounds that these were the cases for which it is most likely that guessing dominated learning in accounting for the TUCE scores observed. The resulting sample contained 178 cases from schools one, two, three, five, and six. The OLS and TSLS regression results are summarized in Table 4.

It is immediately clear that the fit is not very good but that the coefficient of pre-TUCE is significantly positive when estimated by TSLS while it is not when estimated by OLS.¹⁷ Once again, the coefficient of the time variable is estimated to be insignificantly different from zero. It must be remembered that in the model (15) the estimates of the coefficients of the group dummy variables are estimates of logarithmic values.

A test can be formulated of the hypothesis that $\text{Ln}(S_{0i})$ for all $i = 1, 2, 3, 5, 6$.

Let $\hat{\beta}$ = the two-stage least squares estimates of the unconstrained model coefficients (Table 5)

β_0 = the coefficient estimates of the constrained model

$\Sigma_{\hat{\beta}}^{-1}$ = the inverse of the estimated variance covariance matrix of $\hat{\beta}$ from the two stage unconstrained procedure.

Then $(\hat{\beta} - \beta_0)' \Sigma_{\hat{\beta}}^{-1} (\hat{\beta} - \beta_0)$ is distributed as a Chi square with ten degrees of freedom under the null hypothesis.

The χ^2 statistic was estimated to be 3.07 well within the 90-percent acceptance region. Therefore the null hypothesis that there is no difference between control and experimental groups cannot be rejected.

There are at least two kinds of conclusions which can be reached on the basis of the reported research. If the view is taken that the aptitude which enters the learning production function is general aptitude, then the problem of simultaneous equations bias discussed in the first part of the previous section can be solved by using other measures of aptitude than pre-TUCE in the learning equation. No instrumental variables procedure would be needed. If the view is taken, however, that the aptitude which enters the learning production is "aptitude in economics," then the procedures of the previous section would be indicated. In either case, future research in economic learning should strive to collect data which will give information on the aptitude of students in economics to use either as a replacement for pre-TUCE or as an

Table 4
Dependent Variable = $\ln \frac{D \text{ TUCE}}{33 - \text{pre-TUCE}}$

OLS Unconstrained Regression			TSLS Unconstrained Regression		
Variable	Coefficient	Standard Error	Variable	Coefficient	Standard Error
$\ln [T/(1 + T)]$	.48	.73	$\ln [T/(1 + T)]$	.12	.91
$\ln (\text{pre}/33)$	.01	.19	$\ln (\text{pre}/33)$	1.78	.75
GP (0, 1)	-1.75	.27	GP (0, 1)	.01	.79
GP (1, 1)	-1.55	.27	GP (1, 1)	.09	.74
GP (0, 2)	-1.21	.29	GP (0, 2)	.45	.76
GP (1, 2)	-1.32	.28	GP (1, 2)	.49	.81
GP (0, 3)	-1.63	.30	GP (0, 3)	.28	.85
GP (1, 3)	-1.83	.28	GP (1, 3)	-.09	.78
GP (0, 5)	-1.54	.32	GP (0, 5)	.54	.92
GP (1, 5)	-1.50	.30	GP (1, 5)	.54	.90
GP (0, 6)	-.84	.34	GP (0, 6)	1.17	.91
GP (1, 6)	-.94	.31	GP (1, 6)	1.04	.89
$R^2 = .14$			$R^2 = -.31$		
$N = 178$			$N = 178$		

OLS Constrained Regression			TSLS Unconstrained Regression		
Variable	Coefficient	Standard Error	Variable	Coefficient	Standard Error
$\ln [T/(1 + T)]$	.44	.71	$\ln [T/(1 + T)]$	.17	.86
$\ln (\text{pre}/33)$	.01	.19	$\ln (\text{pre}/33)$	1.61	.67
School 1	-1.66	.24	School 1	-.12	.68
School 2	-1.27	.25	School 2	.30	.70
School 3	-1.73	.26	School 3	-.09	.72
School 5	-1.52	.28	School 5	.34	.81
School 6	-.90	.28	School 6	.90	.79
$R^2 = .13$			$R^2 = -.24$		
$N = 178$			$N = 178$		

instrument for it.

Opportunity Cost of Learning

The work of the second and third sections of this article suggests that there is little difference between the control and experimental groups in terms of economic learning. Furthermore, the student study time required to produce this learning appeared to be similar for both control and experimental groups. If student time was of equal value, one could conclude that the added fixed cost of the Martin material (\$350/package, \$150 for duplication, \$3.50/student workbook) is not justified. One caveat to this conclusion is worthy of mention, however.

From the student wage data which was collected on a pre- and postcourse basis, it is clear that not all students faced the same opportunity cost of time. Hourly wages reportedly received by students ranged from zero for unemployed students to over \$9 for students holding part and full-time jobs. Assuming that reported average student wages over the time span of the course reflect the value of time for individual students, then the weekly cost of student study time is simply the product of this wage and their reported average weekly study time. Average aggregate weekly cost figures categorized by change in TUCE points which may be thought of

as average cost of learning data, are given in Table 5. The sample size of 107 for the control and 106 for the experimental group reflects the number of students who were working during the course and reported wage data.

Table 5
Time and Opportunity Cost
of Wage-Earning Students

D TUCE	Average Cost per Week Control	Experimental
- 8	\$24.550	\$ 7.500
- 7	17.962	0
- 6	0	0
- 5	19.816	9.150
- 4	0	0
- 3	23.914	16.234
- 2	36.000	12.647
- 1	15.769	13.013
0	20.811	17.423
1	17.785	13.132
2	17.198	13.407
3	20.162	13.422
4	16.056	14.441
5	14.819	13.152
6	16.610	22.835
7	15.090	13.845
8	24.014	15.520
9	17.234	10.839
10	17.186	11.339
11	28.510	15.249
12	14.394	24.848
13	22.144	18.620
14	0	0
15	0	6.020
16	0	0
17	22.250	0
18	23.520	8.250
	<i>N</i> = 107	<i>N</i> = 106

Casual comparison suggests that the total weekly student cost of learning was less for the experimental group than for the control group, \$1,575.95 per week versus \$1,969.75 per week. This difference may be viewed as sufficient to offset all additional fixed costs of the Martin package; over a ten-week period this would imply nearly a savings of \$4,000.

However, the average weekly student cost of learning per TUCE point, wage $\times$ study time/D TUCE, was not found to be statistically different in simple *t*-test comparisons; *t* = 0.811 for the two groups at the .05 Type I error level.

$$(16) \quad \frac{\text{D TUCE}}{\text{Wage}} = 1.5613 + 0.2155 \text{ Conex}$$

(s.e. = .1914) (s.e. = 2.656)

where Conex = 1 for control and 0 for experimental

The apparent total cost benefit of being in the experimental group may simply be an artifact.

Summary

This research addressed four questions: First, can a learning model be specified within which learning can be justifiably examined in control-experimental groups? Second, is there any difference in the quantity of economics learned between community college control groups and experimental groups using the David Martin audiovisual-tutorial approach? Third, what influence does student classroom and study time have on learning? Fourth, is learning produced in the experimental sections less costly to students than learning produced in the control sections?

The relatively standard learning model in which pre- to postchanges in TUCE scores are regressed on aptitude and other control variables, via OLS, was shown to result in biased and inconsistent estimates when the pre-TUCE is used as a proxy for aptitude. Two stage least squares estimates of the coefficient on pre-TUCE were shown to be significantly positive. In addition, the ceiling effect, known to exist in work with the TUCE, was formally modeled through nonlinear equation specifications for the relationship between aptitude and pre-TUCE and the relationship between changes in TUCE and aptitude. From this equational system the "Gap-Closing Model" was derived as a reduced form equation.

In both the linear and nonlinear model specifications, the hypothesis that increases in economic learning result from the use of the David Martin audiovisual-tutorial package was not accepted. Student classroom and study time also did not show itself to be a significant input in the learning of economics. When properly estimated by TSLS the pre-TUCE effect on student learning was positive and significant. No statistical significant difference was found between the control and experimental groups average weekly student cost of learning per TUCE point.

In short, little difference was found in this study to justify the added cost of the David Martin package. This study does, however, provide a sound statistical modeling procedure which previously has not been attempted in economic education.

Footnotes

¹See, for example [26, 10, 16, 19, 23, 22, 17, 7, 40 and 34]. Walstad [35] provides a review of most of this literature.

²See [33 and 24].

³The importance of model specification in hypothesis-testing has been recognized by statisticians for some time. Educators, such as Dubin and Taveggia [11] and Brown [5] have indicated the need for, and consequence of not, developing formal models describing the teaching-aptitude-learning linkages. Yet, economic education evaluators still carry out simple *t*-test comparisons, assume *ad hoc* linear regression models or include variables as regressors on the basis of a questionable stepwise procedure; see Bishop's 1976 *Delta Pi Epsilon* Research Award winning study [4] as a typical example. Recently, a handful of economic education researchers have come to acknowledge the importance of appropriate modeling; see the Soper-Saunders [30 and 27] interchange and the Soper-Becker-Soper-Highsmith interchange [31, 3, 32, 18].

⁴The problems involved in using the TUCE in community colleges are well documented [9, 36, 21, 37]. Also, see [8] regarding the questionability of using the TUCE in pre-post testing.

⁵Early evaluation by Paden and Moyer [25], Saunders [28], and Lewis and Dahl [20], however, did make use of the average number of hours a student spent studying economics per week. Few other researchers have addressed the impact of time on learning.

⁶The Minnesota Scholastic Aptitude Test Scores (MSAT) or a comparable alternative were available for only 219 students while pre-TUCE scores were available for 330 students. Therefore, using the pre-TUCE as a proxy for aptitude provides more than a 50 percent sample size advantage. In addition, pre-TUCE is a specific measure of economic aptitude while MSAT is a general measure of aptitude. Several authors, however, have been successful in using general aptitude as a determinant of specific economics in a gap-closing model; see, for example [1].

⁷The ceiling effect on the TUCE has been noted by many researchers; see, for example [12, 21].

⁸If satisfactory measures of aptitude are available, researchers will then specify that "value added"—post- minus pretest—is a direct function of the aptitude measure; see, for example [25]. More commonly,

however, an "absolute level model"—posttest is dependent variable—or the "Gap-Closing Model"—change in test scores divided by max-change—is arbitrarily specified with the pretest score and/or aptitude used as a regressor; see, for example [31, 36, 38]. The OLS estimation bias caused by serial correlation inherent in the absolute level model where pre-TUCE is used as a regressor is addressed in [3].

⁹Walstad [35] provides a summary of these yearly reports and gives initial data comparisons.

¹⁰The Minnesota community college instructors had participated in earlier experiments, such as those described in [21 and 38].

¹¹The fact that an individual's characteristics are reflected in his aptitude is recognized by psychologists and educators. For example, Brown [6, p. 314] writes: "An individual's performance on a given task is not determined solely by situation forces *but is also a function of the characteristics of the individual—his aptitudes*" (emphasis ours). This is a controversial issue, however. Therefore, future researchers may want to include explicitly individual characteristics in their learning model specifications.

¹²The presence of bias regressor coefficient estimates also implies that residual estimates are biased. As such, Goldfield and Quandt [15] and Glejser [14] type tests for homoscedasticity are questionable.

¹³See footnotes 8 and 14.

¹⁴Siegfried [29] notes that Whitney [39] was the first to recognize the ceiling effect in economic education test instruments and to propose the "Gap-Closing" measure to account for nonlinearity; see footnote 15.

¹⁵Gery [12] is typically recognized for his work with the TUCE ceiling effects and this "Gap-Closing" measure; see footnote 14.

¹⁶The sample which could be used to estimate the learning equation by TSLS is the subset of the original sample ($N = 330$) containing those 219 cases for which Minnesota Scholastic Aptitude Test (MSAT) scores were available. To test for equal distribution of the 41 negative D TUCE scores throughout the sample for which MSAT scores were available, the sample was divided into quintiles based on MSAT scores. The following table summarizes the number of cases with D TUCE ≤ 0 found in each quintile:

Number of Cases with D TUCE ≤ 0 , by Quintile

Quintile	1st (low)	2nd	3rd	4th	5th
Number of cases with D TUCE ≤ 0	7	14	7	8	5

Under the null hypothesis that the 41 cases are distributed randomly among the 219 observations the equation below is distributed as a Chi-square with four degrees of freedom:

$$\chi^2 = \frac{1}{(0.2)(41)} \sum_{i=1}^6 \left[f_i - (0.2)(41) \right]^2$$

where f_i = number of cases in the i th quintile

The χ^2 value is calculated to be 5.71; the null hypothesis that the cases for which D TUCE ≤ 0 are evenly distributed across aptitude levels is accepted at an 80-percent confidence level.

¹⁷The presence of negative R^2 values for TSLS regressions is the consequence of the manner in which the regression sum of squares is calculated.

REFERENCES

1. Allison, Elisabeth, "Self-Paced Instruction: A Review," *Journal of Economic Education*, 7 (Fall 1975), 4-12.
2. Becker, William E., Jr., Cross-Section Education Evaluation: Source and Results of Serially Correlated Errors, University of Minnesota, Minneapolis, unpublished mimeograph, 1976.
3. Becker, William E., Jr., "Programmed Instruction in Large-Lecture Courses: A Techni-

- cal Comment," *Journal of Economic Education*, 8 (Fall 1976), 38-40.
4. Bishop, Walter L., "Factors Affecting the Level and Development of Economic Understanding of Community College Students," *Delta Pi Epsilon Journal* (ISSN 0011-8052), 1-33.
 5. Brown, Daniel J., "Mirror, Mirror . . . Down with the Linear Model," *American Educational Research Journal*, (Fall 1975), 491-505.
 6. Brown, Frederick G., *Principles of Education and Psychological Testing*. New York: Holt, Rinehart & Winston (The Dryden Press, Inc.), 1970.
 7. Brown, Richard D., "Audio-Tutorial Instruction in Basic Business," *Business Education Forum*, 28 (March 1974), 14-15.
 8. Buros, Oscar Kriesen, *The Seventh Mental Measurement Yearbook, Volume II*. New Jersey: The Gryphon Press, 1972.
 9. Dawson, George G., "The Teaching of Economics in Selected Junior Colleges of Metropolitan New York," New York City Council on Economic Education, 1971.
 10. Doty, Robert B., "Personalized Laboratory Instruction in Microbiology: The Audio-Tutorial Approach," *Educational Technology*, 14 (January 1974), 36-39.
 11. Dubin, Robert, and Thomas Taveggia, *The Teaching-Learning Paradox: A Comparative Analysis of College Teaching Methods*. Eugene, Oregon: Center for the Advanced Study of Educational Administration, 1968.
 12. Gery, Frank W., "Is There a Ceiling Effect to the Test of Understanding in College Economics?" in *Research Papers in Economic Education*, edited by A. Welsh. New York: Joint Council on Economic Education, 1972, pp. 34-49.
 13. Ginman, Peter J., and Robert W. Haseltine, Teaching Basic Macroeconomics via the Audio-Visual-Tutorial Carrel, Center for Economic Education, State University College, Geneseo, New York, unpublished mimeograph, 1972.
 14. Glejser, H., "A New Test for Heteroscedasticity," *Journal of American Statistics Association* (1969), 316-323.
 15. Goldfield, S. M., and R. E. Quandt, "Some Tests for Homoscedasticity," *Journal of American Statistics Association* (1969), pp. 316-323.
 16. Hackett, Dorothy, and Imy V. Holt, "Biological Science as an Audio-Tutorial System of Instruction for the Non-Science Major," *Science Education*, 57 (October 1973), 499-516.
 17. Harriss, Marilyn Francis, The Development, Implementation, Testing and Evaluation of Self-Pacing Auto-Tutorial DOSAGE Forms Laboratory. Doctoral Dissertation, University of Pittsburgh, 1974.
 18. Highsmith, Robert, "Second Generation Research in Economic Education: Comment," *Journal of Economic Education*, 8 (Fall 1976), 48-51.
 19. Howard, Charles, "A Quantitative Appraisal of an A-V-T Program," *Journal of College Science Teaching*, 3 (May 1974), 338-340.
 20. Lewis, Darrell R., and Tor Dahl, "Factors Influencing Performance in the Principles Course Revisited," in *Research Papers in Economic Education*, edited by A. Welsh. New York: Joint Council on Economic Education, 1972, pp. 95-116.
 21. Lewis, Darrell R., Donald Wentworth and Charles C. Orvis, "Economics in the Junior Colleges: Terminal or Transfer?" *Journal of Economic Education*, 4 (Spring 1973), 100-110.
 22. Maccini, John A., Evaluation of an Audio-Tutorial Laboratory Serving College-Level Introductory Geology. Doctoral Dissertation, Ohio State University, 1969.
 23. Marlow, Lorraine Dean, The Development and Assessment of an Audio-Visual Tutorial Laboratory Program Serving College Level Introductory Earth Science. Doctoral Dissertation, Southern Illinois University, 1971.
 24. Martin, David, *Introductory Economic Theory*. New York: McGraw-Hill, 1971.

25. Paden, Donald W., and E. Eugene Moyer, "The Relative Effect of Three Methods of Teaching Principles of Economics," *Journal of Economic Education*, 1 (1969), 33-45.
26. Rowsley, Robert Ellis, A Comparative Study of Two Methods of Instruction in a University Animal Biology Course: Audio-Tutorial with Conventional Lectural Laboratory. Doctoral Dissertation, Auburn University, 1972.
27. Saunders, Phillip, "Needs for Evaluation in Economic Education: A Comment," preliminary paper presented at the National Conference on Needed Research and Development in Pre-College Economic Education, New Orleans, February 1976.
28. Saunders, Phillip, "Student Learning and Instructor Ratings: The Carnegie-Mellon Experience in Introductory Economics," in *Research Papers in Economic Education*, edited by A. Welsh. New York: Joint Council on Economic Education, 1972, pp. 3-11.
29. Siegfried, John J., Is Teaching the Best Way to Learn? An Evaluation of Benefits and Costs to Undergraduate Student Proctors in Elementary Economics, Vanderbilt University, unpublished mimeograph, 1975.
30. Soper, John C., "Needs for Evaluation in Economic Education," preliminary paper presented at the National Conference on Needed Research and Development in Pre-College Economic Education, New Orleans, February 1976.
31. Soper, John C., "Programmed Instruction in Large-Lecture Courses," *Journal of Economic Education*, 4 (Spring 1973), pp. 125-129.
32. Soper, John C., "Second Generation Research in Economic Education: Problems of Specification and Interdependence," *Journal of Economic Education*, 8 (Fall 1976), 40-48.
33. Tolles, N. Arnold, and Peter J. Ginman, Comparative Analysis of an Audio-Visual-Tutorial Process in Basic Macroeconomics, unpublished paper, State University College of New York at Geneseo, 1972.
34. Tolles, N. Arnold, and Peter J. Ginman, "Using an Audiovisual-Tutorial Process in Macroeconomics," *Journal of Economic Education*, 6 (Fall 1974), 57-58.
35. Walstad, William, The Effectiveness of an Audiovisual Tutorial Approach for Teaching Introductory Economics in Selected Two Year Colleges, Master's Thesis, University of Minnesota, Minneapolis, 1975.
36. Weidenaar, Dennis J., and Joe A. Dodson, Jr., "The Effectiveness of Economics Instruction in Two-Year Colleges," *Journal of Economic Education*, 4 (Fall 1972), 5-12.
37. Wentworth, Donald R., The Effectiveness of a Learning Game for Teaching Introductory Economics in Selected Two-Year Colleges. Doctoral Dissertation, University of Minnesota, 1972.
38. Wentworth, Donald R., and Darrell R. Lewis, "An Evaluation of the Use of the Marketplace Game in Junior College Economics," *Journal of Economic Education*, 6 (Spring 1975), 113-119.
39. Whitney, S., "Measuring the Success of the Elementary Course," *American Economic Review*, 50 (March 1960), 159-169.
40. Yeazell, Mary I., A Comparison of Instructor-Paced and Self-Paced Groups in a Criterion Referenced Multi-Media Educational Psychology Course, unpublished paper, West Virginia University, 1974.

Assessing the Impact of an Instructional Innovation on Achievement Differentials: The Case of Computer-Assisted Instruction

John F. Chizmar, L. Dean Hiebert and
Bernard J. McCarney

This paper sets forth an alternative to the conventional procedure for assessing the quantitative impact of an instructional innovation on achievement differentials. The instructional innovation chosen for the purposes of this paper was computer-assisted instruction (CAI.) Any innovation, however, which affects student allocation of time could be analyzed in a similar manner.

The impact of the innovation was evaluated by using a procedure which allows it to affect the production of cognitive achievement in a way which is more general than usually specified. Typically, the effect of an innovation is specified by a neutral shift parameter in the production function for cognitive achievement. In the procedure adopted in this paper, the innovation can also affect cognitive achievement by separately influencing the "productivities" of student characteristics. Using the more general procedure, the differential in mean TUCE scores between users and nonusers of CAI was decomposed into the effect due to CAI and the effect due to differences in student characteristics. Evidence is found that users of CAI performed minimally better than nonusers of CAI as measured by a "TUCE differential," but they did so in spite of CAI, rather than because of it.

The Experiment

The experiment was conducted in the Fall 1975 semester at Illinois State University. A large section ($N = 380$) of the macro principles of economics course constituted the population for the experiment. CAI in economics was viewed as a complement to on-going instruction in that the system was applied as an appendage to a "traditional" macro principles course. The experimental design permits an examination of the impact of CAI within the context of the student's choice decision, since students were given the option to use or not use CAI at their own discretion. In addition, an incentive system was created to encourage student utilization of CAI.¹

In general, a CAI system could increase cognitive achievement for the following reasons:

1. The computer permits elements of self-pacing.
2. The computer provides instantaneous positive reinforcement and instantaneous correction.

John F. Chizmar is Associate Professor of Economics, L. Dean Hiebert is Assistant Professor of Economics, and Bernard J. McCarney is Professor of Economics, Illinois State University.

3. The CAI packages and class attendance may be complementary in the production of cognitive achievement. If the student uses a "review" package after attending the appropriate lecture, the review routine would give him an evaluation of his knowledge, allowing the student to test himself, reinforce concepts presented in the lecture, and remedy deficiencies by means of prompt and further study.

However, the student may perceive that CAI packages can be substituted for class attendance in the production of cognitive achievement. If the use of CAI review packages reduces the cost of production (including time costs), the student may simply adopt the new method in the production of a target level of achievement.³ Hence, one cannot make an *a priori* judgment about the overall impact of CAI on student achievement.

A single output measure was employed to evaluate the CAI technology: student scores on Part I, Form B, of the nationally normed Test of Understanding in College Economics (TUCE), administered in a pre- and postcourse manner.³

Decomposition of the Raw TUCE Differential

If CAI presents a genuine method of enhancing cognitive achievement, then the test performance of the CAI user group should be significantly higher than that of the nonuser group, holding constant other factors such as ability, motivation and maturity which may influence test performance. The conventional format⁴ for assessing the impact of an innovation on cognitive achievement assumes that the sole effect of the innovation is captured by a single coefficient in the equation explaining cognitive achievement. A more general specification of the effect of an innovation would allow all the coefficients to differ between users and nonusers. This allows CAI, say, to affect cognitive achievement by separately affecting the "productivities" of the individual human capital inputs. Because students could select to use CAI, the characteristics of users may be different than the characteristics of nonusers. Thus, cognitive achievement may differ between users and nonusers because characteristics differ and because the coefficients differ. It is of interest, then, quantitatively to assess the magnitude of these two causes of differences in achievement between users and nonusers.

Following a methodology used by Blinder [1] in a different context but which is generally applicable, we propose a decomposition of the raw TUCE differential between users and nonusers into the estimated effects due to differences in individual characteristics and the estimated effects due to differences in coefficients. The latter effect is attributed to CAI. To allow explicitly for differences in coefficients between the two groups, separate regressions were estimated for each group:

$$\text{Post-TUCE}_i^u = \beta_0^u + X_i^u \beta^u + \epsilon_i^u \quad (1)$$

and

$$\text{Post-TUCE}_i^n = \beta_0^n + X_i^n \beta^n + \epsilon_i^n \quad (2)$$

where

X_i = row vector of individual characteristics for the i th student. Includes a proxy for student scores on the post-TUCE (PROXY) designed following a procedure suggested by Soper and Thornton [5], the student's age in years (AGE), student's grade point average (GPA), and student's sex (SEX) measured as a binary variable with 1 = male.

β = column vector of coefficients

ϵ_i = disturbance term

and where the u superscript indicates that the student belongs to the CAI user group and n indicates membership in the nonuser group.

Evaluating equations (1) and (2) at the mean levels for the characteristics and subtracting yields the TUCE differential:

$$D = (\beta_0^u - \beta_0^n) + (\bar{X}^u \beta^n - \bar{X}^n \beta^n) \quad (3)$$

The conventional format assumes that the sole effect of an innovation such as CAI is $(\beta_0^u - \beta_0^n)$. However, using the user group coefficients as weights, the TUCE differential can be decomposed as:

$$D = (\beta_0^u - \beta_0^n) + [(\bar{X}^u - \bar{X}^n) \beta^n + \bar{X}^n (\beta^n - \beta^n)] \quad (4)$$

The first term within brackets is an effect due to differences in individual characteristics and the second term is an *additional* effect due to CAI which has been ignored in the past.

Alternatively, the TUCE differential could be decomposed using nonuser coefficients as weights to yield:

$$D = (\beta_0^u - \beta_0^n) + [(\bar{X}^u - \bar{X}^n) \beta^n + \bar{X}^n (\beta^n - \beta^n)] \quad (5)$$

Since neither decomposition is preferred theoretically, the arithmetic mean of the two procedures will be used.

Table 1 decomposes the TUCE differential into its separate parts. The raw TUCE differential amounts to .6058. The effect of CAI is found by subtracting the effects of differences in individual coefficients. Because there are two weighting schemes, there are two estimates. Using user regression weights, the effect due to CAI equals -.3246. Using nonuser regression weights, the effect due to CAI equals .0873. The arithmetic mean of these two

Table 1
Decomposition of the Raw TUCE Differential

Characteristics	User		Nonuser	
	Regression Weights Adjustment ^a % of D		Regression Weights Adjustment ^a % of D	
D Adjustment for user dif- ferences in:	.6058	100%	.6058	100%
PROXY	.2042	33.71%	.2962	48.89%
GPA	-1.2079	-199.39	-.8994	-148.46
AGE	-.0198	-3.27	.0150	2.48
SEX	.0931	15.37	.0697	11.51
Effect due CAI	-.3246	-53.58%	.0873	14.41%

^aThe user and nonuser regression weights came from estimating equation (1),

$$\text{POST-TUCE}_i^u = 3.159 + .346 \text{ PROXY}_i^u + 3.477 \text{ GPA}_i^u + .186 \text{ AGE}_i^u \\ (.294) (.000) \quad (.000) \quad (.239) \\ + 1.257 \text{ SEX}_i^u \quad (R^2 = .3732) \\ (.007)$$

and equation (2),

$$\text{POST-TUCE}_i^n = 11.503 + .502 \text{ PROXY}_i^n + 2.589 \text{ GPA}_i^n - .141 \text{ AGE}_i^n \\ (.179) (.003) \quad (.000) \quad (.759) \\ + .940 \text{ SEX}_i^n \quad (R^2 = .3731) \\ (.279)$$

The level at which the coefficient is just significant is indicated in parentheses.

estimates is $-.1187$. Thus, CAI accounts for approximately -19.59 percent of the raw TUCE differential.

This analysis, therefore, suggests that the effect of CAI is negative but small. Users as a group performed slightly better in relative terms on the TUCE than nonusers. But they did this *in spite* of CAI primarily because they had more ability as measured by GPA. The analysis also suggests that the students with more ability choose to use CAI, at least within the scenario in which this experiment was run.⁵

These results, however, do not necessarily lead to the conclusion that this application of CAI is cost ineffective. Suppose that students who elect the CAI option are simply substituting CAI review packages for class attendance. If this production technique reduces the costs of production, then (given achievement in economics) the student can produce higher achievement in other courses (or consume more leisure). These latter benefits, when compared with the costs of CAI, could yield a favorable assessment of the cost effectiveness of CAI.

In summary, utilizing a more general specification of the impact of an innovation on cognitive achievement, we find that the effect of CAI on cognitive achievement after adjusting for differences in characteristics between users and nonusers is negative but small. We can only speculate about the reasons for this result. One possible interpretation is that achievement in other courses (together with leisure) is substituted for achievement in economics by students who elect the option to use CAI. An alternative explanation would recognize that the user student must select programs under uncertainty. In the absence of complete information concerning the "payoffs" associated with each program, it is unlikely that users would succeed in finding efficient input combinations to utilize in the production of achievement. The proper interpretation of results concerning the impact of CAI is clearly an important topic for future research.

Footnotes

¹The CAI system consists of three different types of programs that have been labeled "reviews," "demonstrations" and "simulations." This system was developed at the University of Notre Dame by Frank J. Bonello and William I. Davisson [2,3] under a Sloan Foundation grant. The developers view the "review" routines as the heart of this system.

²In fact, Kelley [4] argues that achievement in economics may well be an inferior good.

³The authors chose to use the TUCE as the cognitive evaluation instrument because it will permit comparisons with the experiments conducted by Bonello and Davisson. In addition, the use of the TUCE reduces the possibility of inadvertently preordaining our results by devising an instrument that is biased toward either users or nonusers of CAI. Finally, the TUCE embodies the overall objectives of the macro principles course.

⁴The results obtained from estimating the conventional format are available upon request to the authors.

⁵The conclusion that CAI exerts a significant negative impact on cognitive achievement differs from that reported by Bonello and Davisson. They report [3] on two different experiments which utilize the conventional format. In the first experiment CAI was viewed as a substitute for discussion groups and in the second experiment CAI was used as an appendage to a traditional principles course. The authors analyze both experiments using ordinary least squares regression and conclude from both experiments that CAI had an insignificant effect on cognitive achievement.

REFERENCES

1. Blinder, Alan, "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources* (Fall 1973), 436-455.
2. Bonello, Frank J., and William I. Davisson, "An Integrated System of Computer Assisted Instruction in Economics," paper presented at 1973 Meetings of the Midwest Economics Association.
3. Davisson, William I., and Frank J. Bonello, *Computer Assisted Instruction in Economic*

Education (University of Notre Dame Press, 1976).

4. Kelley, Allen C. "The Students as a Utility Maximizer," *Journal of Economic Education*, 6 (Spring 1975), 82-92.
5. Soper, John C., and Richard M. Thornton. "Self-Paced Economics Instruction: A Large Scale Disaggregated Evaluation," *Journal of Economic Education*, 7 (Spring 1976), 81-91.

What do Economics Majors Learn?

By DAVID G. HARTMAN*

The introductory economics course has been studied extensively, with respect to both course content and instructional methods. Far less is known about the remainder of the program for undergraduate economics majors. As an initial step toward understanding and improving the educational experience of economics majors, this paper examines what economics majors learn. Specifically, the primary results give information about what students at Harvard learn, but it is hoped that insights into the experience of students in general can be gained. Because my primary concern is with general economic understanding, the emphasis of this paper will be on how well students learn micro- and macroeconomic concepts and how to apply them. The study begins with a few casual observations.

Before any formal analysis was undertaken, there were reasons to believe that improvements in the economics program were needed. First, as one who has taught introductory economics at Harvard and also been in charge of the general examinations required of all graduating economics majors, I have often suspected that either students in the introductory course know much less than it appears or there is a sizable group of economics majors who leave school with little more than the level of economic understanding they achieved by the end of their first course. This evidence is, of course, of limited usefulness because the passage of time would tend to erode the students' skills. Moreover, most graduating

seniors are at least a year away from their last general micro or macro course. A more formal analysis will be presented below to isolate the impacts of various parts of the economics program on student performance.

A secondary source of casual evidence is the students themselves. It is no secret that a large majority of Harvard economics undergraduates feel that the introductory course is the high point of the program, with a number of the subsequent courses, particularly the intermediate theory sequence, considered highly repetitious of material covered in introductory economics. The feeling is widespread that a general analytical ability is not being developed as it should be in an undergraduate program. There are several points to be considered about such student complaints. First, having witnessed similar students' feelings at another school (where the point was not made as vocally as at Harvard), I find it difficult to accept this dissatisfaction as unique. Whether it highlights a significant problem is a separate issue. Although I tend to take fairly seriously the complaint of a broad group of students that courses are not sufficiently rigorous, it is certainly possible that a great deal of useful skill is being acquired. It may be that some courses only seem repetitious because the beginning course provides at least a brief introduction to all the major topics to be examined. Graduate teaching assistants in the intermediate micro-macro sequence support this hypothesis by pointing out that students often perform poorly because, having the illusion of already knowing the material, they put too little effort into the intermediate courses. Finally, it is far from clear, without further investigation, that repetition is not a valuable use of time, even if students complain.

*Harvard University. I wish to thank President Derek C. Bok for his interest in and funding of this effort. I owe Elizabeth Allison a great debt for her encouragement and advice. Liam P. Ebrill, Jeff Wolcowitz, and Ken Sokoloff did a heroic job of extracting the needed data from general examinations. David Lindauer provided many useful comments on an earlier draft.

TABLE 1—CHARACTERISTICS OF HONORS AND NONHONORS ECONOMICS STUDENTS

	Honors Students	Nonhonors Students	All Students
Number in entire 1977 class	61	118	179
Number of students in sample	50	44	94
Number of students having taken micro courses:			
Traditional micro course	43	26	69
Policy micro course	4	5	9
Graduate micro course	3	2	5
Number of students having taken no micro course	0	11	11
Number of students having taken macro courses:			
Traditional macro course	27	18	45
Graduate macro course	23	10	33
Number of students having taken no macro course	0	16	16
Average introductory micro course grade (0-15 scale)	11.45	9.98	10.79
Average introductory macro course grade (0-15 scale)	11.86	10.56	11.27

These issues are essentially empirical; empirical evidence to be presented will hopefully provide some answers. The investigation will proceed by first discussing specific goals of the economics program, then suggesting a method for measuring how well students meet these goals, and finally attempting to discover the contributions of various parts of the program to the students' achievement.

The general goals of an economics program are highly controversial, as Robert Horton and Dennis Weidenaar discovered in their survey. The more specific objectives implicit in the Harvard method of evaluating graduating students seem roughly consistent with the Horton and Weidenaar "consensus" goal; the Harvard general examinations in economics are designed to test a student's ability to use economic tools to answer real world questions. Specifically, a student's knowledge of macro theory, knowledge of micro theory, ability to apply macro theory, and ability to apply micro theory are evaluated and can be given separate scores. For purposes of this paper, attainment of skill in those four categories will be taken as the objective of the program. Because it represents the best information available,

the score on the general exam in each category will be used to measure what an economics student has learned. The problems associated with using test scores to represent achievement should be kept in mind as the empirical results are analyzed.

1. Majoring in Economics at Harvard

Economics students choose between an "honors" major and a "nonhonors" major. In a typical year, there are about 75 honors and 125 nonhonors majors. Only honors students are required to complete a course in both intermediate micro and macro and to write a senior thesis. Although nonhonors students have neither requirement, many complete the intermediate theory courses because of an interest in the topic, because they wish to be well prepared for the general exam, or because they had at some time intended to pursue an honors program. A good grade record is required for graduation with honors, so there is a natural presumption that honors majors are better students. However, there are a number of reasons for good students to do a nonhonors program, including the desire to take a wider range of courses than the honors requirements would allow. Table 1

gives comparisons of 1977 honors and nonhonors students in terms both of grades in introductory courses and of elective courses taken. The comparisons are based on the sample of students to be used in the empirical study.¹

Honors students have a number of options in meeting the intermediate micro-macro requirement. The courses with the largest numbers of students are the "traditional" micro and macro offerings. The empirical investigation which follows is based on students who took the 1977 general examinations, but who took the traditional micro or macro course at a variety of points in their careers, with different instructors and texts. In addition, the traditional courses apparently are becoming more like the "alternative" intermediate courses than were the traditional courses taken by most 1977 graduates.

There are two other courses in micro which satisfy the honors requirement. Almost 10 percent of the sample students took the policy and applications oriented micro course, which is taught by the case method with a large number of problem sets and a great deal of assigned reading. The other course, with about 5 percent of the sample students, is taught at nearly the level of sophistication of the course intended for economics Ph.D. students. It requires a mathematical preparation beyond most undergraduates and is also taken by graduate students from other departments. Because of this variety in the Harvard program, it will be possible to determine the impact on learning of the traditional micro course compared to two very different alternatives (to be called the "policy" micro course and the "graduate" micro course); a reference group is available since 12 percent of the students had had no micro course past the introductory level.

There is only one alternative to the tradi-

tional macro offering. It was taken by 35 percent of the sample students. Despite the fact that this course is intended to be taught at nearly the level of the course for economics Ph.D. students, and is taken by graduate students from other departments, the level of math required is not above that of most Harvard undergraduates. The material presented not only is more difficult than that in the traditional macro course (particularly since the course employs an extensive reading list rather than a basic text), but also puts more emphasis on policy. An attempt will be made to assess the impacts of the traditional course and this graduate course on student performance. Once again, the reference group consists of those students (17 percent of the sample) who had had neither course.

II. Measurement and Empirical Results

Graders of the 1977 general examinations were asked to assign each student four grades (each on a 0-10 scale): 1) micro theory; 2) micro application; 3) macro theory; and 4) macro application. Each question was graded by two graduate students familiar with the undergraduate micro-macro courses. Since different questions are asked on the honors and nonhonors exams, it was important that specific grading standards be adopted. The graders reported little problem with consistency in scoring the two exams because of the similarity and generality of the questions asked. However, a concern was expressed, particularly by the macro graders, about their ability to separate the students' knowledge of theory from their ability to apply it, based on answers to these general questions. The empirical results tend to confirm this difficulty with macro; I suspect that it is in the nature of macroeconomics for theory and application to be so closely related.

To allow a direct confirmation of grading consistency across the two examinations, a "practice examination" was given to thirty of the students two weeks prior to the general exams. The practice test consisted of multiple choice questions chosen to fall into the four categories of interest. By hav-

¹The 94-student sample on which these comparisons are based results from requiring that personal data be available and that the students included took the spring 1977 exams. Nonhonors students may substitute a fall exam, so they are underrepresented in the sample. I have no reason to believe, however, that this sample is biased in any way critical to the results.

ing a set of scores on a common test to compare with the scores on the honors and nonhonors exams, it was possible to verify that in no case was there any significant bias in the scoring of honors and nonhonors general exams.

Separate regressions were run to estimate the impact of courses taken as part of an economics major on students' measured knowledge of micro theory, ability to apply micro theory, knowledge of macro theory, and ability to apply macro theory. A simple model is assumed: an economics major's level of skill in each of the four categories depends on his/her innate ability, the skill developed in introductory economics, and the amount learned in the relevant courses taken as part of the economics program. Measurement of the first two factors will be discussed as the empirical results are presented. The contribution of the economics program to a student's skill, obviously the factor most important to isolate for purposes of this paper, is estimated using a set of variables indicating which courses a student has taken. With respect to microeconomics, three dummy variables indicate whether a student has taken the traditional micro course, the policy micro course, or the graduate micro course (or, of course, none of these). Another variable is the number of "micro related field" courses a student has taken.² In the macroeconomics regressions, two dummy variables indicate whether a student has taken the traditional macro course or the graduate macro course (or neither). Another variable is the number of macro related field courses a student has taken.³

In the initial attempt to explain students'

knowledge of micro theory (equation (1) in Table 2), grades in the microeconomics half of the introductory course were used to represent both ability and the learning acquired in the course. The rest of the equation consists of the "economics major" variables discussed above. The dependent variables and introductory course grade variables are expressed in *logs* in every estimation. So, the coefficients on the micro course dummy variables and the field course variable in (1) are the estimated percentage increases in micro theory score produced by taking the associated course. Therefore, the traditional course is estimated to increase one's micro theory score by about 24 percent, the policy micro course by 44 percent, the graduate micro course by 51 percent, and an additional micro field course by 9 percent. All of these effects are significant at or very close to significance at the 5 percent level.

These results should, however, be regarded with suspicion because of inclusion of only the introductory course grade to measure ability. In particular, the equation (1) result could occur as a consequence of students choosing courses based on their ability. For example, if the most able economics students choose the graduate micro course, its larger estimated impact need not be evidence of any learning premium over the traditional course. I would anticipate that the most significant bias results from not having any measure of math skill in equation (1), since the best information a student has about his aptitude for micro (aside from math) is probably his introductory course grade.

Numerous measures of student ability including SAT scores, high school class rank (adjusted for school size), and interview ratings, as well as data on race and sex, were available from admissions information. The SAT math scores, as expected, were the only significant ability factor in any regression. The inclusion of the SAT scores (equation (2)) reduces to insignificance the introductory micro course grade variable. Since the sign is then wrong, the coefficient is constrained to be zero.

²As used in this study, microeconomics related field courses are defined as: economic principles and public policy; public finance; development economics; international trade and investment; economics of managerial decisions; business organization and behavior; markets and market structure; labor economics; urban economics.

³As used in this study, macroeconomics related field courses are defined as: monetary theory and financial institutions; applied macroeconomics; public finance; international monetary economics; development economics.

The resulting equation (3) has substantially lower estimated impacts for components of the economics program, confirming the extent of bias introduced when math ability is left out. Now, only the policy micro course meets the strict criterion for significance normally employed; the number of micro field courses does, however, have a significant impact at the .05 level in a one-tail test. The graduate micro course (.10 significance level) and the traditional course (.20 level) do not fare so well. The small number of students having taken the graduate course (see Table 1) may be a contributing factor to its insignificance.

The corresponding equation ((6)) for ability to apply micro concepts indicates that, of the components of the economics major, only the policy micro course has an impact on student performance significant at even the .10 level. The traditional course and the field courses produce an estimated improvement in a student's ability to apply microeconomics with a significance of .20; the influence of the graduate micro course is even less significant.

The results explaining students' knowledge of macro theory and their ability

to apply it lend support to the contention that it is difficult to separate the two scores; the results are very similar. The results also show that math ability does not play a significant independent role.

Table 2 shows the significance level of the graduate macro course in explaining the macro theory score on the general examinations is .001, while the traditional course and the macro field courses are significant at the .05 level. Taking the graduate macro course produces an estimated 21 percent increase in exam score, while the impact of the traditional course is 11 percent and the effect of each macro field course is an estimated 4 percent.

The large differences in the macro theory and applications scores seem to imply that the traditional course and the field courses have a less significant effect on students' ability to apply macro concepts. It should also be noted that the R^2 of the applications equation is substantially below that of the macro theory equation. This implies that the ability to apply macro concepts is more difficult to measure and/or is determined more by the ability factors not controlled for, than is knowledge of macro theory.

TABLE 2

Equation	Dependent Variable (lags)	Introductory Course Grade (lags)		Traditional Intermediate Course (dummy)		Policy Intermediate Course (dummy)		Graduate Intermediate Course (dummy)		Field Courses Taken (number)		SAT Math Score	SAT Verbal Score	R^2
		Constant	Micro	Micro	Macro	Micro	Micro	Micro	Macro	Micro	Macro			
(1)	Micro Theory Score	1.11 (4.12)	.088 (.83)	.239 (1.92)		.438 (2.47)	.510 (2.36)		.094 (2.69)					.14
(2)		-7.21 (3.39)	-.063 (.60)	.129 (1.09)		.360 (2.16)	.302 (1.46)		.056 (1.67)			1.263 (2.82)	.096 (.25)	.31
(3)		-6.91 (3.36)		.126 (1.07)		.360 (2.17)	.300 (1.46)		.057 (1.71)			1.177 (2.79)	.112 (.30)	.31
(4)	Micro Application Score	1.31 (3.92)	-.018 (.14)	.303 (1.97)		.375 (1.71)	.384 (1.44)		.096 (2.22)					.06
(5)		-9.26 (3.58)	-.222 (1.73)	.168 (1.17)		.294 (1.44)	.119 (.47)		.047 (1.15)			1.859 (3.41)	-.133 (.29)	.27
(6)		-8.20 (3.22)		.154 (1.06)		.294 (1.43)	.114 (.45)		.050 (1.21)			1.559 (2.98)	-.077 (.17)	.24
(7)	Macro Theory Score	1.34 (7.03)	.174 (2.13)		.107 (1.90)			.218 (3.32)		.039 (1.78)				.37
(8)		1.22 (1.07)		.173 (1.98)		.107 (1.87)		.214 (3.03)		.040 (1.80)		.103 (.46)	-.085 (.43)	.38
(9)	Macro Application Score	1.58 (6.91)	.073 (.74)		.086 (1.27)			.231 (2.93)		.040 (1.54)				.28
(10)		.78 (.57)	.054 (.51)		.081 (1.17)			.213 (2.52)		.042 (1.58)		.129 (.48)	.002 (.01)	.29

Note: Absolute values of t -statistics are shown in parentheses. To obtain all data, sample was restricted to 94 students. Therefore, significance levels (one-tail test) are approximately: $t = 2.37$, significance level = .01; $t = 1.99$, significance level = .025; $t = 1.66$, significance level = .05; $t = 1.29$, significance level = .10.

III. Summary and Conclusions

There are numerous reasons for statistically insignificant results and, therefore, there is great risk in placing too much emphasis on the empirical evidence presented here. However, those concerned with what economics majors learn can take little comfort from this analysis.

The traditional microeconomics course appears to have a small impact on either students' knowledge of micro theory or their ability to apply it to real world problems at the end of their college careers. One of the alternatives offered to Harvard students, a graduate-type highly theoretical course, does a bit better in adding to the students' knowledge of micro theory, but is worse when it comes to teaching them to apply the theory. Only the policy oriented micro course has a verifiable effect on the students' understanding of micro theory; even so, its impact on their ability to apply micro concepts is below usually acceptable significance levels, once adjustments are made for mathematical ability.

The traditional intermediate macro course gives evidence of improving macro theory scores, although its effect on ability to apply those concepts is below usual significance standards. The graduate level macro course, which is oriented to policy and not particularly mathematical although demanding, is a highly significant explainer of student knowledge. Finally, the micro and macro field courses tend overall to have a moderate impact on student scores.

At the risk of overemphasizing these admittedly rough statistical conclusions, it appears that there is a problem with the economics major. After years of study and improvement in the beginning course, it seems time to begin a similar effort with respect to the rest of the undergraduate program. It is not surprising that with a beginning course which has become a

thorough and rigorous introduction to the discipline, students find the traditional courses which follow repetitious. They may be mistaken and, in any event, repetition may be quite valuable, but the evidence available does not substantially dispute their conclusions. Study and innovation, which are taking place in traditional courses at Harvard in response to the changes in the knowledge possessed by students emerging from the introductory class, must be encouraged. At a minimum, new methods of presenting information should be devised to prevent student dissatisfaction. From Harvard's experience with alternative courses, it would appear that at least the average student is capable of learning significantly more difficult material, by reading more extensively and working on more independent assignments, than is normally taught in intermediate courses. On the other hand, it seems unproductive to offer courses too much like those given for graduate students because, without experience in problem solving, undergraduates do not learn to apply the theory they are taught.

As indicated at the start, this paper represents just an initial step. The conclusions are necessarily imprecise and must be sharpened by study of systems at other schools. Until such broader information is available, application of these results to other than the specific courses studied here is quite risky. The effects on student learning of courses other than those in micro and macro and the impact of the economics program on knowledge retained years after graduation are also areas in which further research is necessary.

REFERENCES

- R. V. Horton and D. J. Weidenaar, "Wherefore Economic Education?," *J. Econ. Educ.*, Fall 1975, 7, 40-44.

ECONOMIC EDUCATION

TIPS and Technical Change in
Classroom Instruction

By ALLEN C. KELLEY*

This paper presents some research results on the efficiency of a teaching technique (*TIPS*) I have used in the Principles of Economics course at the University of Wisconsin-Madison. My presentation will be extremely selective; only the results of five out of some ten different output measures are reported. I shall argue that *TIPS* represents an improvement over the commonly employed lecture-discussion classroom technology. The model I use to evaluate the efficiency of *TIPS* is broader than that usually used in educational research, since it takes into account not only the total magnitude of the benefits and costs, but also their distribution.

TIPS

TIPS (Teaching Information and Processing System) is a testing and evaluation system which provides the capability of increasing the level of individualized instruction in the classroom (Kelley, 1968, 1970). *TIPS* enables the instructor to prepare, administer and process short multiple-choice "surveys" on a regular

basis throughout the semester. Based on the results of each survey and on instructions or "decision rules" provided by the professor, a series of instructional reports are prepared and printed by data processing equipment. Under normal circumstances the surveys are given once every week and require ten to fifteen minutes of class time. To date the surveys have not been used for grading; they have been administered to provide interim information used to diagnose student difficulties and to prescribe remedies before major examinations take place.

Three major sets of instructional reports are generated by *TIPS*. A *Student Report*, prepared for each student in the class, is available three to four hours after the student leaves the classroom. This report provides individualized assignments for each student based on his measured proficiency on the various concepts covered on the *TIPS* survey. A student performing well on one concept may receive an enrichment and/or optional assignment; on another concept, where deficiency is revealed, he may receive a lower-level required assignment. Assignments may also be based on the student's learning skills, e.g., his mathematical versus his verbal ability. The *TIPS* survey results permit identifying well before formal examinations those students who are failing the course; individual tutorials and compensatory instruction may then be arranged. Superior stu-

* Professor of Economics, University of Wisconsin-Madison. This research was supported by the Esso Education Foundation. The Graduate School of the University of Wisconsin provided substantial computer time. The assistance of James K. Matson, Robert M. Schmidt, and Michael L. Wiseman should be conspicuously acknowledged. I have also benefited from the comments of Rendigs Feis, W. Lee Hansen, and Burton A. Weisbrod. Given space limitations, the present paper has been severely condensed. An expanded version is available from the author upon request.

ECONOMIC EDUCATION

dents may be provided the option of writing papers or engaging in research in lieu of taking the exam.

Summary reports are prepared for both the professor (covering class performance) and the teaching assistant (covering the results of each TA's sections). These reports provide the feedback required by the instructor to modify his class assignments and presentations in response to revealed deficiencies and strengths. The TA reports also provide a list of assignments required of each student, together with lists of students who, for example, are required to establish tutorials or who have been provided the option of writing papers and engaging in term-paper research.

TIPS employs some of the oldest principles of instruction but uses modern technology to provide each student a course of instruction appropriate to his needs. The degree of individualized instruction facilitated by *TIPS* is largely invariant to class size. This approach is applicable to a wide range of disciplines where course objectives can in large part be measured by well-designed objective-type questions. The system is designed to attenuate instructional problems inherent in higher education where expanding enrollments and rising costs are accompanied by large class sizes and where student abilities and interests span a wide spectrum.

A Model of Educational Evaluation

The model of educational evaluation employed below emphasizes the *distribution*, as well as the total magnitude, of benefits and costs associated with alternative instructional approaches (Hansen *et al.*). The pervasive failure to consider distributional issues in educational evaluation is tantamount to assuming either that students are a homogeneous group—each student receives the same amount of outputs and sustains the same amount of costs, and the outputs and costs are valued

equally for each student—or that students should be treated as if they were a homogeneous group. Two implications of these observations relate to the appraisal of the literature evaluating educational technology. First, there is an abundance of studies which fail to identify a statistically significant impact on student performance of alternative educational approaches; this may, as McKeachie has observed, result from the fact that “methods optimal for some students are detrimental to the achievement of others” (p. 1157). Second, if students do benefit differentially from alternative teaching techniques, the statistical models which omit these distributional effects are misspecified; they produce statistically biased and typically uninterpretable results.

The Experimental Evaluation of *TIPS*

In the fall of 1968 *TIPS* was utilized in an experiment in the Principles of Economics course at the University of Wisconsin-Madison. The objective was to assess the impact of *TIPS* on student achievement, student attitudes toward the course and *TIPS*, and student retention of economic principles (measured one year later). Students in the experimental group employed *TIPS* during the first eight weeks of the course. Students in the control group were taught using a lecture-T. A.-homework assignment format thought representative of that widely employed in the principles course. The control and experimental student groups met with their professor three times a week; the fourth session, a discussion led by a graduate teaching assistant, met in smaller groups of fifteen to twenty-five students. The total amount of homework assignments in the two groups was approximately the same. In the control group all students received an identical “average” assignment. In the experimental group the students in difficulty received larger and lower-level as-

AMERICAN ECONOMIC ASSOCIATION

signments; those demonstrating proficiency were given optional, ungraded assignments.

The control and experimental lectures, each comprising about 250 students, met with the same instructor at contiguous lecture hours in different buildings. In both groups students received almost identical lectures. *TA's* were randomly selected from the Departmental "pool" and were assigned to one of the two lectures without bias. Identical text materials were required. Subsequent analysis reveals that a minimum of student interaction between the two groups occurred; furthermore, statistical tests show that the two lecture groups possessed a statistically identical distribution of attributes: aptitude, prior academic achievement, sex, academic major, class, and mathematics background. The "Hawthorne effect," likely present, was attenuated by the procedure of briefing students in the *TIPS* class with the use of a nonpromotional, printed document describing the system and the experiment. This avoided any tendency toward overemphasis of the experiment by the instructor.

Considerable care was taken to obtain output measures which were valid and unbiased. A two-hour mid-term examination contained twenty multiple-choice questions drawn from those provided in the instructor's manual to the text; none of the questions had appeared on *TIPS* surveys. The student also answered five

short-answer questions of an applied-problem type, and had a choice of one of two long essays. The short-answer questions and essays were equally divided between questions submitted by *TA's* in the two lectures. Students in both lectures were administered identical examinations at the same hour (different buildings). Elaborate precautions were taken to ensure objectivity of grading: multiple choice questions were machine graded; the remaining portions of the tests were anonymized by removing student names and assigning a numerical code for subsequent reassembly. All responses to a particular question by students from both lectures were graded by a single *TA*. Undoubtedly the grading possessed significant variance in terms of accuracy; we assert, however, that there was no bias which would preclude an objective appraisal of the impact of *TIPS*.

Impact of *TIPS* on Student Achievement

Space precludes the presentation of the theoretical model underlying the analysis, the results for each of the output measures, a discussion of the estimation format employed, and a defense and interpretation of each variable included in the regression. The results presented in equation (1) for the aggregate score from the first midterm examination are representative of the output measures over that portion of the course when *TIPS* was used. Least squares regression procedures were employed; *t* statistics are in parentheses.

$$\begin{aligned}
 (1) \quad 0 = & 18.35 + .17 \text{ ACTSAT} + .08 \text{ LogHSP} + 3.23 \text{ SOPH} + 3.95 \text{ UPPER} + .31 \text{ MATH} \\
 & (5.18) \quad (5.98) \quad (2.85) \quad (3.08) \quad (2.59) \quad (.28) \\
 & + .94 \text{ PSENG} + 1.84 \text{ BUS} + 1.16 \text{ ECON} + .71 \text{ COMASGN} + .09 \text{ PCTSECT} \\
 & (.67) \quad (1.63) \quad (.62) \quad (3.11) \quad (2.98) \\
 & + 1.56 \text{ ASNDONE} - .30 \text{ LogASNDONE} \cdot \text{ACTSAT} + 38.35 \text{ TIPS} - 3.96 \text{ ASNDONE} \cdot \text{TIPS} \\
 & (2.17) \quad (-1.66) \quad (3.26) \quad (-2.15) \quad \text{TIPS} \\
 & + .82 \text{ ASNDONE} \cdot \text{LogACTSAT} \cdot \text{TIPS} - 6.39 \text{ LogACTSAT} \cdot \text{TIPS} \quad \bar{0} = 52.09 \\
 & (1.77) \quad (-2.20) \quad r^2 = .34
 \end{aligned}$$

ECONOMIC EDUCATION

Student achievement was positively and significantly related to the number of homework assignments completed (*ASNDONE*), the percentage of sections attended (*PCTSECT*), whether the student was a sophomore or an upperclassman (*SOPH*, *UPPER*), his *ACT* or *SAT* score, his graduating high school percentile ranking (*HSP*), the difference between the number of his required assignments and those handed in (*COMASGN*—a measure of commitment?), and whether he was in the *TIPS* class. Neither knowledge of calculus (*MATH*) nor major (*PSENG*, *BUS*, *ECON*—physical science or engineering, business, economics) contributed significantly to examination achievement.

An interpretation of the results in (1) is facilitated by comparing the aggregate performance of four "representative" students. Table 1 presents the predicted score of four students and the percentage contribution to that score due to the several independent variables. Charles Kinbote and Jack Gradus are "average" students in the *TIPS* and control groups, respectively; they are "twins" in all respects except *TIPS*. John Shade is a relatively low achiever in the *TIPS* class; Sybil Swallow is relatively bright.

As in most studies, prior aptitude and achievement constituted the most important independent variables, accounting for around 25-35 percent of the explained variance. Section attendance counts sig-

nificantly and positively, although these results do not necessarily measure the absolute contribution of the *TA*'s or sections (i.e., there was no control), but rather the impact of differential section attendance, a measure which could be a proxy for student attributes such as study discipline, interest in the course, and so forth.

Homework assignments were most beneficial to the relatively slow student as measured by *ACT-SAT* scores; they were of little significance for the bright student. This result illustrates one way in which *TIPS* could possibly increase the efficiency in the use of instructional inputs and, in this case, the student's time. Since bright students in the *TIPS* class received very few required assignments, *TIPS* was likely instrumental in increasing the productivity of instructional resources. Not only did bright students "save" ten to fifteen hours per semester by not working assignments of low productivity, but *TA*'s spent no time grading and recording the results. Instructional resources were instead shifted toward the low-achieving student, where the relative productivity of the homework-assignment technique appears to be high.

The contribution of *TIPS* to student examination performance was greatest for the relatively low-achieving student (19.5 percent), falling to 13.3 percent for his brighter classmate. The impact of *TIPS* occurs not only through individualized homework assignments, but also through

TABLE 1—PERCENTAGE CONTRIBUTION OF EACH FACTOR TO THE PERFORMANCE OF INDIVIDUAL STUDENTS ON THE FIRST MIDTERM EXAM

Name of Student	Predicted Score	Intercept	Class	MATH	HSP	ACT-SAT	PCT-SECT	Major	COMASGN	ASNDONE + ASN, LogACT	TIPS + Interactions*
John Shade (Slow Student <i>TIPS</i> class)	54.63	33.6%	5.9%	...	10.0%	5.7%	14.4%	3.4%	-0.0%	7.6%	19.5%
Charles Kinbote (Average student <i>TIPS</i> class)	58.36	31.2%	5.5%	...	11.0%	16.9%	13.3%	3.1%	-0.0%	3.1%	15.8%
Sybil Swallow (Bright Student <i>TIPS</i> class)	63.69	28.8%	5.1%	...	11.4%	24.7%	12.3%	2.9%	-0.0%	1.5%	13.3%
Jack Gradus (Average Student Control Class)	49.51	37.1%	6.5%	...	13.0%	20.1%	15.9%	3.7%	-0.0%	3.7%	

* Interactions = $TIPS \cdot ASNDONE + TIPS \cdot ASN \cdot \text{Log}(ACT) + TIPS \cdot \text{Log}(ACT)$.

feedback and study discipline. All of these influences are plausibly more beneficial to the low achiever. Moreover, given the experimental design, it is not surprising that the low-achieving student received greater benefits from *TIPS*. A larger share of instructional resources (grading and *TA*'s time) was allocated to this student, even though the *total* resources employed in each class (including the student's time) was roughly the same.

Are these measured impacts of *TIPS* specific to the "type" of examination question? The results in Table 2 suggest that

TABLE 2—CONTRIBUTION OF *TIPS* TO STUDENT SCORE ON THE COMPONENTS OF THE FIRST EXAMINATION

Student	Multiple Choice	Short-Answer	First Essay
John Shade (low achiever)	16.2%	29.4%	16.5%
Charles Kinbote (average achiever)	13.3%	27.0%	9.7%
Sybil Swallow (high achiever)	10.8%	24.3%	6.6%

the greatest impact of *TIPS* was revealed on short-answer, applied problem-type questions; moreover, the low-achieving student gained as much on the essay question as he did on the multiple-choice questions. Finally, the distributional impact of *TIPS* appears consistently across all types of questions, although it is most pronounced on the essay.

Other Effects of *TIPS*

1. Student Evaluation of *TIPS*

Students responded favorably to the employment of *TIPS*. They possessed no significant hostility to the use of data processing equipment. Moreover, they urged that *TIPS* be used in future economics classes and in other disciplines as well. The student's evaluation of *TIPS* is largely invariant to his class, major, or *ACTSAT* standing (Kelley, 1968).

2. Student Evaluation of the Course and the Professor

The student's evaluation of the course and the professor was uninfluenced by *TIPS*. The end-of-course evaluations (prepared by the Department of Economics and by the Wisconsin Student Union) yielded virtually identical results in the control and the experimental groups. This evidence is consistent with the hypothesis that the Hawthorne effect was unimportant.

3. The Lasting Effects

Approximately 250 students were retested one year after the completion of the experiment. While the results are not yet completely analyzed, preliminary findings reveal that the differential *TIPS* effect is maintained over time, although it diminishes somewhat in magnitude. This impact of *TIPS* on the retention of knowledge is probably attributed to the change in study habits engendered by the teaching approach. Students in the *TIPS* class have been shown to study and review continuously throughout the semester, allocating a relatively smaller share of their time to preparing for major examinations (Kelley, 1968, pp. 451–52). This contrasts with the "typical" study pattern of allocating a greater proportion of time to examination preparation, i.e., cramming. The latter study pattern has been shown by psychologists to represent a relatively unproductive format if knowledge retention is the criterion of evaluation.

The most interesting finding on the retention of knowledge is that the distributional effect of *TIPS* largely disappears. If this result stands up to further analysis, then clearly the "efficiency" assessments made above regarding the relative productivity of allocating a disproportionate share of instructional resources to the lower-achieving student could well be modified, and even overturned.

4. The Number of Majors

The proportion of the *TIPS* class selecting economics as a major, as measured two years later, was 23 percent higher than that in the control class. This unexpected result is somewhat difficult to interpret. Recall that students appeared to obtain no differential "enjoyment" of the course or instructor due to *TIPS*. Possibly their greater academic success in the course, by comparison with their evaluation of it, is the more important factor in their selection of a major.¹

Cost of *TIPS*

Costs can be divided into six categories: 1) physical facilities, 2) data processing, 3) faculty time, 4) student time, 5) *TA* time, and 6) other (secretarial, administrative, printing, and so forth). A detailed examination of the differential *total* costs reveals that there is no significant difference between the per student cost in the *TIPS* and the control lectures. This somewhat surprising result is obtained from the fact that the increased direct costs of the system (computer time, professor's time in survey preparation, printing) is largely offset by the more efficient use of existing classroom resources (*TA* grading time). If an evaluation of the student's time "saved" or released by *TIPS* is taken into account, then *TIPS*, as implemented during the experiment, would have economized on total instructional resources.

Research to date has not yet determined the effects of *TIPS* on the *distribution* of costs. The major distributional impact occurs in the allocation of *TA* and student time. To the extent that *TIPS* is not used for enrichment purposes, then students of

lower achievement are, on the one hand, incurring greater time costs and, on the other hand, receiving a disproportionate share of the benefits and instructional resources.

TIPS and Economic Efficiency

It is possible to form a preliminary appraisal of the efficiency of *TIPS* as used in this particular experiment. *TIPS* produces increased output for most students although, as implemented, more output was distributed to the relatively low-achieving student. Assuming a positive value of marginal output, then the sign of the total value of output is positive and is uninfluenced by the distributional effects of this instructional technique. However, *TIPS*'s distributional impact influences the *size* of the value of total benefits.

While the total cost of *TIPS* is approximately the same as in the traditional classroom format, a higher cost was assumed by the low-achieving student. Assuming that the opportunity cost of the time of this student is less than or equal to that of his brighter counterpart, I can conclude that *TIPS* is a more efficient technique than the traditional classroom framework.

Several qualifications are in order. First, these conclusions are based largely on the course examination measures. Other measures, including output-added, measures of intellectual curiosity, or critical thinking, may yield quite different findings. Second, the value of the output depends on *who* is doing the valuing. While faculty may be inclined to value strongly the impact of *TIPS* on retained student achievement of economics, students, in contrast, plausibly place a relatively high weight on course "enjoyment," somehow measured. (We have concluded that course "enjoyment" is largely invariant to *TIPS* use.) Alternatively, even the most enlightened departmental chairman, while responsive to achievement and course evaluations, will

¹ The same percentage distribution of A's, B's, and C's was awarded to each class. A difference in letter grades in this range did not therefore account for the larger number of majors from the *TIPS* class. This grading procedure was considered necessary to ensure student cooperation during the experiment.

AMERICAN ECONOMIC ASSOCIATION

place some positive (relatively high?) weight on the "economics-majors" output. Finally, my results apply to one experiment, with one instructor, in a single university, and in a particular course. Even if we assume that the experiment is methodologically sound, and that analytically sensible theoretical and statistical models were employed, the ability to generalize from this single experiment is limited. We would be interested in replicating this experiment in other courses, disciplines, and environments. Moreover, these experiments should ideally be outside the direct influence of the researcher.

A final qualification relates to the predicted outcome of replicating *TIPS* in other settings. A wide variation of *TIPS* impacts is likely to be identified. The impact of *TIPS* is in large part a reflection of the relative success with which the professor correctly selects the appropriate teaching instruments and test items. Given the paucity of scientifically based findings on the relative productivity of alternative teaching approaches and materials, the full potential of a *TIPS*-like approach to

instruction will not be revealed until major advances are made in the more fundamental instructional areas of testing, diagnosis, and prescription. If the past can be taken as a rough guide to the future, notable advances in these areas of instruction, in economics or any other discipline, are not likely to be just around the corner.

REFERENCES

- W. L. Hansen, A. C. Kelley, and B. A. Weisbrod, "Economic Efficiency and the Distribution of Benefits from College Instruction," *Amer. Econ. Rev.*, May 1970, 60, 364-69.
- A. C. Kelley, "An Experiment with *TIPS*: A Computer-Aided Instructional System for Undergraduate Education," *Amer. Econ. Rev.*, May 1968, 58, 446-57.
- , "The Economics of Teaching: The Role of *TIPS*," in K. G. Lumsden, ed., *Recent Research in Economics Education*, Englewood Cliffs 1970, 44-66.
- W. J. McKeachie, "Research on Teaching at the College and University Level," in N. L. Gage, ed., *Handbook of Research on Teaching*, Chicago 1963.

A Training System for Graduate Student Instructors of Introductory Economics at the University of Minnesota

Darrell R. Lewis and Charles C. Orvis

College and university teaching is . . . the only profession (except the proverbially oldest in the world) for which no training is given or required.
—Jacques Barzun, 1968.

This paper reports on an evaluation of the effectiveness of a three-part training system developed at the University of Minnesota for assisting graduate student instructors (GSI) of the introductory course in economics.

Historically, the University of Minnesota (along with most other large universities) has made extensive use of graduate students in the teaching of undergraduate economics. There are no prerequisites for teaching the principles course other than being enrolled as a second year graduate student in economics and being eligible for financial aid. Until three years ago, the department simply provided a syllabus, a textbook with instructor's manual and section assignments with room numbers, and turned the graduate student instructors loose in the classroom.

However, during the past three years an integrated series of student evaluations, videotaped classroom observations and instructional seminars have been developed for training and assisting the 22 graduate students who are providing the instruction for 44 sections of the introductory course at the University of Minnesota.

In the fall of 1970, in response to a request from the Graduate Economics Club as well as the Economics Department's own desire to improve instruction, a series of departmental seminars on the teaching of economics for all new GSIs was initiated. Simultaneously, the possibility of videotaping individual instructors was announced by the University's Radio and Television Department. Since this seemed to offer unexplored possibilities for the improvement of instruction, the project was integrated into the seminar concept. Approximately 20 GSIs and senior faculty were videotaped, critiqued and reviewed by selected members of the department. Some of the tapes were also used as part of the seminar—to show others the benefits of videotaping and to demonstrate various teaching techniques. A student evaluation questionnaire (with specific instructor performance criteria identified) was also administered to each GSI's class to provide additional feedback to each instructor as to his performance.

Throughout the first year both participation and feedback from the GSIs had been excellent. However, a basic question remained: Were we having any measurable impact on both student and instructor performances in the classroom? To resolve this question,

Darrell R. Lewis is Professor of Economic Education and Charles C. Orvis is a Research Fellow at the University of Minnesota. The authors wish to acknowledge the invaluable assistance in the developmental stages of this project by John Scarbrough and Ray Riezman, Research Fellows at the University of Minnesota.

the following study was conducted.¹

Experimental Design

During the 1971 fall quarter, all students enrolled in Economics 1-001 (Principles of Economics—Macroeconomics) were selected as a control population. This total population was divided into fourteen sections which met three times each week as a section and once a week for a mass lecture. Enrollment in each of the sections was essentially a self-selection process on a first-come, first-serve basis. Students were not aware of which staff member would be assigned to any of the sections being offered between 8:00 a.m. and 4:00 p.m., Monday through Friday. The average size of the sections was 25. The mass lecture was handled by senior faculty while the fourteen sections were conducted by seven graduate student instructors (GSI), each teaching two sections.

During the fall Quarter, the seven GSIs were precluded from participating in or having knowledge about the videotaping or seminar. Similarly, these seven instructors for the control groups and their students were unaware of both the experimental design and the hypotheses being tested. However, all of the Economics 1-001 students in the fall term responded to questionnaires dealing with student characteristics and were pre- and posttested on the Test of Understanding in College Economics (Part I, Forms A and B). Postcourse student evaluations of each instructor's performance were also collected on the Purdue Rating Scale for College Instructors.²

The Purdue Rating Scale for College Instructors is a recently developed semantic differential questionnaire with 27 items. Each question associates with one of five factors representing the instructor's (1) personal characteristics, (2) objectivity, (2) exposition, (4) tests and grades, and (5) subject matter knowledge. Similarly, each question is posed in such a fashion as to give description for appropriate corrective action—i.e., each is expressed in performance criterion terms. Reliability tests and appropriate validation for the instrument have been produced at Purdue University.

In order to control for the experimental training of instructors, the *same seven GSIs* were used as the experimental group during the winter quarter when Economics 1-001 was again offered. The experimental group of 438 winter quarter students was again divided into fourteen sections with an average section size of 31. As with the control students, all the winter quarter experimental students responded to the questionnaire, the Purdue Rating Scale, and the TUCE. Subsequent tests on selected student characteristics and pre-TUCE scores revealed no significant differences between the control and experimental groups (see Table 1 and study results below). All sections and instructors in both the fall and winter quarters used the same instructional materials, senior faculty for the mass lectures, and departmental course syllabus.

The experiment was designed in such a way that the seven instructors were randomly selected from a total of 22 GSIs in the fall of 1971. The seven instructors were then given only a syllabus and section assignments and were not provided with any other assistance or training—i.e., the norm for most GSIs at large universities. However, during the winter quarter these same seven GSIs were systematically exposed to the department's three-part training system.³

The GSI Training System⁴

Each week throughout the winter quarter the seven experimental instructors met together in an informal seminar with the authors of this study and another senior faculty member from the economics department. Such topics as the purpose and scope of introductory economics, student-teacher interaction and discussion techniques, teaching techniques for various concepts, integration of supplemental readings and lectures with the syllabus and text, orientation and familiarization with the literature on teaching at the college level, introduction to the economic education literature at the collegiate level, how to plan and establish learning objectives for each class or unit, and how to construct tests and measure student performance made up the content of the seminar.

As a second component of the GSI training system, each instructor was videotaped three times during the quarter. Following each 45-minute videotaped class session, approximately two hours were spent reviewing and critiquing each tape with the individual instructor.⁵

In conjunction with the videotaping, two instruments were developed to assist both the instructor and the reviewing procedure.⁶ Prior to each class designated for videotaping, the instructor completed a questionnaire directed to the objectives, content and techniques expected to be covered during the class period. This information was subsequently reviewed and compared with the videotape during the critiquing session.

The second instrument was constructed so as to measure actual instructor performance from the videotape. Prior to the review session with each instructor, a specially trained graduate student in economics previewed and coded each tape at twenty-second intervals according to a specially adapted observation scheme which measured (a) the method employed (lecture, question/problems, discussion, other), (b) the learning objectives (knowledge of facts, theoretical concepts, exposition on theory, simple application, complex application) and (c) the verbal and nonverbal expressions (supportive, receptive, neutral, unresponsive, disapproval). The data were then summarized and presented to the GSI during the review session.⁷

The data from this latter instrument have proven valuable in at least two regards. Instructors, like students, respond to those things which are being measured. Secondly, the instrumentation was able measurably to confirm or reject those things the instructor said (thought) he was doing in his classroom. It also reinforced the reviewer's intuitive critique and comments.

The third facet of the training system involved student evaluations of the instructors' performance. As discussed above, the Purdue Rating Scale for College Instructors was given to all students in each of the seven GSIs fall classes. The results were then discussed during the winter quarter videotaping review sessions with each of the instructors. Suggestions and strategies for improvement for each low-rated item were then developed. This instrument and this procedure facilitated not only the training of GSIs, but provided for built-in evaluative comparisons with the subsequent winter quarter instructor ratings.

Description of Experimental Results

As Table 1 indicates, the winter quarter experimental group did not differ significantly from the fall quarter control group in any of five matching variables—i.e., Sex, Age, Cumulative Grade Point Average, ACT Score, and Pre-TUCE—at the two-tailed .05 criterion level being employed in this study. Consequently, with the same instructors teaching in both the fall and winter quarters, the groups were considered adequately matched for the purposes of this study.

The Pre-TUCE data in Table 1 also indicate that the Minnesota scores for both the experimental and control groups approximate the national norm of 13.24 at the outset of each quarter term. Post-TUCE scores for the fall quarter control group also approximate the national norm of 19.08, further indicating normality for the control sections [8].

Impact of the Training System on Student Learning

As Table 1 indicates, the winter quarter experimental students clearly outperformed the control students in economic understanding. Not only were the differences between group Post-TUCE scores significant, but the Change-in-TUCE scores (Post-TUCE minus Pre-TUCE) also indicated significant differences. The experimental group exhibited a 54 percent gain over their Pre-TUCE score while the control group experienced only a 43 percent gain in output added.⁸ The gains for the control group are comparable to the national norming data for the TUCE wherein students from four-year colleges showed average gains of 40.3 percent. The experimental group's performances were clearly superior.

Table 1
Description of Student Characteristics, Performances and Evaluations:
Fall and Winter Quarters

Variables	Fall Quarter N = 323 14 Sections		Winter Quarter N = 438 14 Sections		t-test Comparing Means
	Means	S.D.	Means	S.D.	
Sex (0, 1) Male = 1	.71	.45	.70	.46	.30
Age (17-38)	20.49	3.09	20.80	2.74	1.43
Grade point average (0-4)	2.73	.51	2.77	.48	1.09
ACT score (0-36)	24.88	3.18	24.46	3.63	1.50
Pre-TUCE (0-33)	13.52	3.70	13.04	3.96	1.71
Post-TUCE (0-33)	19.46	4.70	20.11	4.53	1.97†
Change-in-TUCE	5.94	4.52	7.07	4.67	3.36††
Average instructor rating) (1-6), 1 = very low	4.11	.84	4.46	.78	5.68††
Rating scale subparts:					
1. Personal evaluation (1-6), 1 = very low	4.41	.90	4.76	.85	5.42††
2. Objectivity evaluation (1-6), 1 = very low	4.32	.91	4.70	.86	5.82††
3. Exposition evaluation (1-6), 1 = very low	3.82	1.10	4.32	.99	6.45††
4. Testing evaluation (1-6), 1 = very low	3.89	.95	4.10	.92	3.05††
5. Knowledge evaluation (1-6), 1 = very low	4.26	.93	4.52	.81	4.02††

† Significant at the .05 level.

†† Significant at the .01 level.

Although the significance of the training system's impact on student learning is clearly evident from the above-given data and discussion, a more controlled analysis can be performed by fitting the student descriptors, evaluations and test results to a multiple linear regression model.

The equation for the model takes the familiar linear form wherein $Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_7X_7 + e$. Post-TUCE, the dependent variable, is assumed to be linearly related to the following independent predictor variables:

X_1 = Pre-TUCE (0 - 33), continuous

X_2 = ACT score (0 - 36), continuous

X_3 = Cumulative grade point average (0 - 4), continuous

X_4 = Age (17 - 38), continuous

X₅ = Sex (1 = male, 0 = female), continuous

X₆ = Instructor evaluation (1 - 6; 1 = very low), continuous

X₇ = Class type (1 = experimental in winter quarter, 0 = control), dichotomous

Five of the variables (Pre-TUCE, ACT, GPA, Age and Sex) were chosen on the basis of past research, generalizability and their qualities as educational realities which teachers of the introductory course must accept (and can measure) when trying to influence student learning.⁹ The last two variables included in the model (Instructor evaluations and Class type) were unique to the situation studied.

This type of class attended (experimental or control) was the key variable in this study. In the multiple linear regression model of this study, the coefficient of this variable measures the residual contribution of the GSI training system to student achievement in introductory economics.

A stepwise regression procedure was employed which entered each variable into the equation in order of significance. The data, as fitted to the multiple linear regression model described above, are presented in Table 2.

Table 2
Regression Results from GSI Training Program
(*t*-statistic in parentheses)

Independent Variables	Dependent Variable: Post-TUCE	
	<i>N</i> = 761 Students (323 Control and 438 Experimental)	
X ₁ Pre-TUCE (0-33)	.34	(7.44)††
X ₂ ACT (0-36)	.36	(6.81)††
X ₃ GPA (0-4)	.02	(4.70)††
X ₄ Age (17-38)	.19	(3.25)††
X ₅ Sex (0-1; 1 = male)	1.43	(3.88)††
X ₆ Instructor evaluation (1-6; 1 = Very Low)	.43	(2.16)†
X ₇ Class type (0-1; 1 = experimental in winter quarter)	.71	(2.12)†
Constant	-5.24	
Standard error	3.88	
Adjusted R ²	.30	

† Significant at .05 level

†† Significant at .01 level

As Table 2 indicates, when the data were fitted to the multiple linear regression model, the significance of the earlier *t*-statistics was confirmed. Controlling for prior knowledge in economics (X₁), mental ability and achievement (X₂, X₃), maturation (X₄), sex (X₅) and student evaluations of the instructor (X₆), the type of class (X₇) with experimental involvement in the project did have a significant association with the students' Post-TUCE scores.¹⁰ The model predicts that a student attending a class which was involved with the GSI training system would, on the average, score almost three-quarters of one point (.71) more than nonparticipants on their Post-TUCE scores.

The regression model also indicates that the six other variables significantly associate with student achievement in economic understanding. Prior knowledge in economics (X₁), mental ability and achievement (X₂, X₃), maturation (X₄), sex (X₅) and student evaluations of the instructor (X₆) were all found to be significant.¹¹ These findings are all consistent with the results of other research in this field [6].

Impact of the Training System on Instructor Performances

The data in Tables 1 and 2 also confirm that the GSI training system had measurable and significant influences on the instructors' actual performances as measured by student evaluations. Not only was there a significant difference between quarters in the total Rating Scale for all of the instructors, but each of the subparts to the Rating Scale were also significantly different between the experimental and control groups. In turn, these significant changes in GSI performances associated significantly with student learning as confirmed by the Instructor evaluation variable (X_4) in the regression model of Table 2 with Post-TUCE.

It is important to note that throughout the experimental quarter's videotaping review sessions each instructor was presented with the Rating Scale (student evaluations) results from the previous course. Suggestions and strategies for improvement were then developed with each instructor for each low rated item. The instrument and these procedures were apparently effective.¹²

Individual instructor ratings are summarized in Table 3 for each of the two groups with respect to the Rating Scale and its subparts. With one major exception (Instructor IV), all the GSIs increased their rating scores for almost all the subparts to the Rating Scale. It is important to note that the only instructor whose ratings dropped (Instructor IV) developed mononucleosis during the experimental quarter and was the least active and enthusiastic participant in the training system. This illness and behavior undoubtedly carried over into his teaching performance as well. In testing, for example, he simply pulled old exams from his files. It is also interesting to note that Instructor V was an office mate with Instructor IV and used the same tests as did Instructor IV. Consequently, both instructors went down in their student ratings dealing with "tests and grades." Both the students and the Rating Scale instrument are apparently sensitive to such behavior and circumstances.

The student evaluations, as revealed by the Rating Scale, were also substantiated in early videotape reviews during the experimental quarter. Both the reviewer in his observations and the actual videotape coding procedure revealed the same strengths and weaknesses as the student evaluations of GSI performances. High instructor ratings on "Personal characteristics" and "Exposition" skills were supported by high coding frequencies on "Supportive" and "Receptive" categories of verbal and nonverbal expressions; high instructor ratings on "Subject matter knowledge" were supported by high coding frequencies on teaching methods other than "Lecture" and on higher level learning objectives such as "Complex applications." The consistencies between these two instruments, along with the actual videotaped observations, were persuasive evidence in getting the GSIs to change their teaching behavior.

Summary

This study has confirmed that a systematic teacher training program involving Graduate Student Instructors of introductory economics with an integrated series of student evaluations, videotaped observations and instructional seminars can have a significant and measurable impact on both student and instructor performances in the classroom. Specifically, it was found that as a result of the training system (a) student performance, as measured by the TUCE, and (b) instructor ratings, as measured by the Purdue Rating Scale for College Instructors, both increased significantly. It was also found that instructor ratings, as measured by student evaluations on the Rating Scale, associate highly with student performances on the TUCE.

The experimental efforts and results of this study suggest that other institutions and departments of economics can and should undertake greater responsibilities for providing their graduate student instructors with teacher training. The specific components of the Minnesota training system are not costly in either set-up terms or in maintenance. Most of the developmental costs have already been assumed in the crea-

Table 3
Student Ratings of Instructors

Subgroups		Range	Average: All Instructors	Individual Instructors						
				I	II	III	IV	V	VI	VII
Personal characteristics (Questions 2-3-6-17-21-25)	Fall	3.90-4.97	4.39	3.90	4.97	4.09	4.89	4.65	4.18	4.03
	Wtr.	4.33-5.41	4.74	4.33	5.41	4.55	4.54	4.78	4.69	4.86
	Change from fall to winter		.35	.43	.44	.46	(.35)	.13	.51	.83
Objectivity (Questions 7-10-19-23)	Fall	3.90-4.91	4.30	3.98	4.91	4.19	4.59	4.60	3.95	3.90
	Wtr.	4.26-5.26	4.68	4.46	5.26	4.56	4.26	4.80	4.59	4.83
	Change from fall to winter		.38	.48	.35	.37	(.33)	.20	.64	.93
Exposition (Questions 11-13-14-15-22)	Fall	2.86-4.59	3.78	3.47	4.59	3.70	4.49	4.12	3.20	2.86
	Wtr.	3.82-5.14	4.30	3.82	5.14	4.16	4.16	4.37	4.17	4.25
	Change from fall to winter		.52	.35	.55	.46	(.33)	.25	.97	1.39
Tests and grades (Questions 1-4-9-16-20-27)	Fall	3.18-4.50	3.87	3.34	4.32	3.63	4.50	4.25	3.84	3.18
	Wtr.	3.52-4.62	4.08	3.52	4.62	4.28	3.97	4.12	4.10	3.97
	Change from fall to winter		.21	.18	.30	.65	(.53)	(.13)	.26	.79
Subject matter knowledge (Questions 5-8-12-18-24-26)	Fall	3.28-4.89	4.24	4.23	4.69	4.15	4.89	4.32	4.09	3.28
	Wtr.	4.17-4.98	4.52	4.55	4.98	4.42	4.44	4.46	4.62	4.17
	Change from fall to winter		.28	.32	.29	.27	(.45)	.14	.53	.89
Overall (Averaging above subgroups)	Fall	3.45-4.70	4.11	3.78	4.70	3.95	4.67	4.39	3.85	3.45
	Wtr.	4.14-5.08	4.46	4.14	5.08	4.39	4.27	4.51	4.44	4.42
	Change from fall to winter		.35	.36	.38	.44	(.40)	.12	.59	.97

tion of the instruments for evaluation and codification. The fixed costs for the videotape equipment totals only \$2,400. Only three real departmental resource inputs are needed for maintaining and monitoring the system: (1) the senior faculty member meeting with the seminar for approximately 10-15 hours each quarter, (2) an undergraduate student with a quarter-time appointment for setting up, recording, and taking down the videotape equipment for each class session recorded, and (3) an outstanding graduate student instructor with a half-time appointment in economics for previewing, coding and critiquing each class session tape and for handling the logistics of the seminar, student evaluations and videotaping procedures. Each of the two students can be trained for the system in less than five hours. With these resources, experience at Minnesota indicates that approximately 7-10 GSIs can be processed through the entire system within one quarter—i.e., the seminar, three videotaping episodes for each instructor and student evaluations. The Department of Economics at Minnesota has found the results in student and GSI performances and GSI satisfaction to be worth the costs in time, effort and dollars.

Footnotes

¹The study described in this paper was conducted experimentally during 1971-72 and was fully implemented during 1972-73.

²Copies of the Purdue Rating Scale for College Instructors as well as the survey questionnaire will be sent by the authors on receipt of a self-addressed envelope with sufficient postage for 2 ounces.

³Although most of the GSIs in this experimental study had teaching experience prior to their participation in the project, a weakness in the design of the study is the possibility that any superior performance in the winter quarter may be attributed to their maturation and/or additional experience. However, similar data on student and GSI performances from an earlier study with the introductory course at the University of Minnesota indicate that the additional experience on only one term is not significant [5]. In fact, the obverse was true in the current study. The GSI who had the most previous teaching experience was the GSI in the fall term with the lowest student and instructor performances and subsequently showed the most improvement as a result of the training system. On the other hand, the two GSIs with the least teaching experience were among the top three in student and instructor performances during the fall term.

⁴A number of excellent discussions concerning the reliability, validity and usefulness of many of the procedures and components included in the Minnesota GSI Training System can be found in recent publications by Eble [2,3,4], Costin, *et al.* [1], and Nowlis, *et al.* [7].

⁵As additional incentive for soliciting GSI participation in the project, the "best" instructors were encouraged to retain the videotape of their best performance (usually their third) and include this as a part of their vitae for subsequent employment. Additionally, each of the participating instructors was assured of anonymity throughout the entire training and evaluation proceedings. Both of these practices have continued in the subsequent training of GSIs at the University of Minnesota.

⁶Copies of both instruments will be sent by the authors on request.

⁷An additional "Summary Checklist" for the videotape reviewing was also constructed and will be sent by the authors on request.

⁸Any discussion of output added on the TUCE must be qualified with the recognition that the output-added function is clearly nonlinear; there are easy questions, questions of "medium" difficulty, and some which are very difficult. In fact, the test was designed this way in terms of cognitive composition. It is therefore somewhat inappropriate to compare increments on this test, as constructed. For example, at the extreme a student moving his total score from 3 to 6 on the TUCE has picked up much less economics than a student moving from 28 to 31. Only on a truly "linear test" can these types of comparisons be safely made.

⁹A number of other possible independent variables were considered for inclusion. However, such other variables as math background, major, and family background were found to be nonsignificant in other similar studies and/or intercorrelated with those identified in this study. Moreover, for policy purposes only those independent variables which were identifiable prior to the course were included.

¹⁰All variables in this model were found to have intercorrelations of .21 or less in the correlation matrix except ACT and Pre-TUCE. They had a correlation of .31, a degree of intercorrelation but not detrimental to the model's analysis since they were both significantly correlated with Post-TUCE.

¹¹The possibility of significant interaction terms (between X_4 and X_7 , X_2 and X_7 , X_5 and X_7 , X_2 and X_4) was also examined in subsequent regression models. The results were essentially the same as those found in Table 2. No significance was found in any of the added variables.

¹²The reliability, validity and usefulness of student ratings of college teaching are also persuasively presented in an excellent review article by Costin, Greenough and Menges [1].

REFERENCES

1. Costin, F., W. T. Greenough and R. J. Menges, "Student Ratings of College Teaching: Reliability, Validity, and Usefulness," *Review of Educational Research*, 41 (December 1971), 511-535. (Summarized elsewhere in this Journal.)
2. Eble, K. E., *Career Development of the Effective College Teacher*. Project to Improve College Teaching. Washington, D.C.: American Association of University Professors and Association of American Colleges, 1971.
3. Eble, K. E., *Professors as Teachers*. San Francisco: Jossey-Bass Inc., 1972.
4. Eble, K. E., *The Recognition and Evaluation of Teaching*. Project to Improve College Teaching. Washington, D.C.: American Association of University Professors and Association of American Colleges, 1970.
5. Lewis, D. R., and T. Dahl, "Factors Influencing Performance in the Principles Course Revisited," in *Research Papers in Economic Education*, Arthur Welsh, Ed. New York: Joint Council on Economic Education, 1972.
6. Lewis, D. R., and C. C. Orvis, *Research in Economic Education*. New York: Joint Council on Economic Education, 1971.
7. Nowlis, V. C., K. E. Clark and M. Rock, *The Graduate Student as Teacher*. Washington, D.C.: American Council on Education, 1968.
8. The Psychological Corporation, *Manual Test of Understanding in College Economics*. New York: The Psychological Corporation, 1968.

ECONOMIC EDUCATION: ASPIRATIONS AND ACHIEVEMENTS

By G. L. BACH AND PHILLIP SAUNDERS*

In recent years, widespread concern has been expressed over the low level of economic literacy of the general public. Since apparently only a small portion (perhaps 10-20 per cent) of all future citizen-voters will ever take as much as one economics course in college,¹ attention has been increasingly focused on the possibility of developing at least a modicum of economic understanding through teaching in the high schools.

Concern for this problem has developed gradually within the economics profession, culminating in the policy statement adopted by the AEA's Executive Committee in March, 1964 and the establishment of a new action-oriented Committee on Economic Education [9, p. 565]. Prior to this action the Association's previous Committee on Economic Education had for some years attempted to stimulate economists to help high school teachers and administrators with this task, mainly through attention to the problem at annual AEA meetings. During the past five years, however, a more concerted effort has developed, including at least four major projects, two of which have involved AEA participation or sponsorship.

This article summarizes briefly these recent developments and then presents the findings of recent investigations of what economics is actually being taught in the high schools and by whom, and of the effec-

*The authors are, respectively, professor and assistant professor of economics at Carnegie Institute of Technology. The former is chairman of the AEA Committee on Education, and this article represents, in part, a report to the profession on the results of several special undertakings the Association has sponsored or encouraged. The authors wish to thank especially their colleagues, Michael Lovell, Lester Lave, and Leonard Rapping, for statistical advice at a number of points; and Mrs. Ann Brunswick of the National Opinion Research Center who supervised the NORC study on which a substantial portion of this report is based. NORC has a separate detailed report on its findings which includes extensive information on high school teachers and their backgrounds not reported here [2].

¹Less than 50 per cent of all present high school students will enter college, and apparently only about one-fourth, certainly less than one-half, of these will take a course in economics.

tiveness of recent AEA-sponsored steps to improve economic understanding. The major sections are: I. Recent Steps to Improve Economic Understanding; II. What Economics Do Students and Teachers Know?; III. Current Materials and Teaching Practices; IV. The Teachers; V. Who Watched "The American Economy"?; VI. Effectiveness of "The American Economy"; and VII. Some Implications. For those who want a quick overview, Tables 1 and 6, plus the concluding "Implications," may be helpful.

I. Recent Steps to Improve Economic Understanding

In 1959 the AEA Committee on Economic Education appointed a 13-man Textbook Study Committee to analyze the economic content of the textbooks being used in high school social studies courses (in economics, problems of democracy, and American history). Paul Olson served as the chairman. This committee reported to the profession in "Economics in the Schools," published as a supplement to the *American Economic Review* in 1963 [11].

In 1960 the Association appointed a National Task Force on Economic Education, an independent group of well-known economists, to describe for high school administrators and teachers a minimum core of economic understanding fundamental to good citizenship and reasonably attainable by most high school students. The need for such a statement from the profession had been widely voiced by school teachers and administrators, school boards, and leading citizens. The report of the Task Force, *Economic Education in the Schools*, was published in 1961 [12]. Some 250,000 copies of this report have been distributed in full or in summary form to laymen and to secondary school administrators and teachers throughout the United States.²

In 1961 the AEA agreed to serve as co-sponsor of a new year-long national television course on economics, called "The American Economy," and offered in "The College of the Air" series carried by 182 CBS stations and virtually all the educational television stations in the United States during 1962-63 (and rebroadcast by many in 1963-64).

²Members of the Task Force were G. L. Bach, Chairman, Lester Chandler, R. A. Gordon, Ben Lewis, and Paul Samuelson from the profession; Arno Bellack and M. L. Frankel from secondary education; and Floyd Bond, Executive Secretary. The report is available from the Committee for Economic Development, 711 Fifth Ave., New York 22, N.Y., which provided funds to finance the Task Force's work and published the report, with no control over, or responsibility for, the conclusions reached. The CED also played a leading role in stimulation action on "The American Economy" and the "Test of Economic Understanding" described below, and in securing funds to finance these projects, but in each case with no control over the project itself or the findings of the economists and other professionals involved. The CED's policy statement, *Economic Literacy for Americans* [13], provides a summary of its role, and its own position on some of the substantive issues involved.

Professor John R. Coleman of Carnegie Institute of Technology was the national teacher; some 40 leading economists participated as guest lecturers; and the members of the National Task Force, at the request of the AEA, served as a policy and advisory committee in planning the series. The course was aimed to help the general public, students, and especially school teachers with inadequate backgrounds in economics. It was based substantially on the report of the National Task Force as to content and approach.³ "The American Economy" had an average daily audience of over one million viewers (according to a standard Nielsen survey and reports from the participating educational television stations), including some 15,000 high school teachers and about 5,200 students and teachers who registered for college credit at one of the 361 cooperating colleges and universities. The cost of the program was approximately \$1.5 million, which was provided by The Ford Foundation and a large number of leading business firms; in addition CBS contributed free air time. This course, and the AEA sponsorship of it with the Joint Council on Economic Education and Learning Resources Institute (the producer), followed the general patterns successfully used by the professional associations in physics, chemistry, and mathematics in preceding years.

Following completion of "The American Economy," the National Task Force and the Learning Resources Institute commissioned the National Opinion Research Center, affiliated with the University of Chicago, to conduct a major national study of what economics is being taught in the high schools and by whom, and to analyze the effectiveness of "The American Economy" in raising the level of economic understanding among high school social studies teachers, one prime target of the program. The study was outlined by the National Task Force and conducted by NORC on the basis of a national stratified cluster sample. A cross section of high schools in the United States was selected. Teachers in those schools were questioned and they took a special test on economic understanding. The 2,791 actual questionnaire responses were then weighted to represent 4,677 high school social studies teachers, out of an estimated total of approximately 60,000-65,000 such teachers in the nation. The teachers thus surveyed completed detailed questionnaires on their backgrounds, their teaching practices and attitudes, and the content of their courses. In addition, each teacher took a 25-question version of the nationally standardized "Test of Economic Understanding" described below. All questionnaires and tests were administered by NORC personnel. The results of this study

³ A course outline, including the names of participating economists, was reported in the *AER* [10] and is available in more detail in the special guide prepared for the course [4]. Professor Coleman has also presented his philosophy of the course in [3].

provide the first detailed picture, based on a scientifically designed sample of all high school social studies teachers, of who the teachers of economics are in the high schools, what their backgrounds are, what they are now teaching, and how much they learned from "The American Economy."

Simultaneously, in 1962-64, a new standardized high school level "Test of Economic Understanding" was developed by a related group of leading economists, educational psychologists, and high school educators to help school administrators and teachers evaluate how much economics their students know and are learning, using a large national norm group for comparison. This step, financed by the CED through the Joint Council on Economic Education, but entirely under the control of the experts involved, was an attempt to meet the widespread demand for such a test that would be professionally competent and free from the suspicion of bias often leveled against tests devised by representatives of business or other economic groups.⁴ The test committee, under the chairmanship of Dr. John Stalnaker, President of the National Merit Scholarship Corporation, spent approximately a year devising a "Test of Economic Understanding" for use with high school students at the 9th to 12th grade level, using the report of the National Task Force on Economic Education as a rough guide as to the concepts and areas to be tested. The test was designed for students with or without separate courses in economics, and dual forms were prepared so the test could be given to students on a before-and-after basis. Because the test was designed for use by thousands, even hundreds of thousands, of students in widely varying schools and areas, the use of "objective" type questions was mandatory. All questions are multiple choice, with four alternative answers given to each.

Questions cover the basic areas of micro- and macroeconomics, plus some questions on "applied" fields such as international economics and comparative systems, but they omit all technical detail beyond such simple concepts as supply and demand. There are a few factual questions, but most are focused on the understanding of basic "concepts" and ability to handle "problem" or "application" situations. A few questions call for analysis of major policy issues (such as monetary and fiscal policy), and three require reading and interpreting graphs of economic time series.⁵ Roughly, one-third of each test covers micro-

⁴The economists on this committee were G. L. Bach (Carnegie Tech), E. O. Edwards (Rice), J. A. Kershaw (Williams), Ben Lewis (Oberlin), and Lewis Wagner (University of Iowa).

⁵Typical questions are:

Factual:

1. In large business corporations common stockholders generally do not:
 - a. own the business

BACH AND SAUNDERS: ECONOMIC EDUCATION

economics, one-third macroeconomics, and one-third special topics and applications, such as the farm problem, international trade, comparative economic systems, and the like. Each question was pretested on thousands of students, many were extensively revised, and the test was carefully balanced for coverage, concepts involved, degree of difficulty, and kinds of understanding—all within the practical limits imposed by the need for a mass-testing instrument. The final test is extremely elementary. Every economist would probably devise a somewhat different one for the purpose; the constraints imposed by test experts and the realities of mass use in high schools were considerable.

Beyond these major recent moves, culminating in the AEA policy statement in June, 1964 [9], many other groups have worked steadily to improve economics in the schools. The nonpartisan Joint Council on Economic Education has perhaps worked most closely with professional economists, and the AEA has long nominated economists to be members of the JCEE Board of Trustees. Important though these efforts may have been, however, it is impossible to detail them here.

-
- b. receive a share of the profits
 - c. vote for the board of directors
 - d. manage the day-to-day business

Concept Understanding:

1. Under a private enterprise economy the function of competition is to:
 - a. eliminate wasteful advertising
 - b. eliminate interest and profits
 - c. prevent large firms from driving small ones out of business
 - d. force prices to the lowest level consistent with a reasonable profit
2. A rise in the price of which product would be likely to increase the demand for butter?
 - a. butter
 - b. oleomargarine
 - c. bread
 - d. any of the above

Problem Analysis:

1. If, when there is full employment, the federal government increases its spending without increasing its tax revenues, generally:
 - a. a serious depression will occur
 - b. an increase in unemployment will occur
 - c. the national debt will decrease
 - d. inflation will occur
2. In a basically private enterprise economy, which tax is likely to alter most the pattern of consumer choices among alternative products?
 - a. a general sales tax
 - b. a personal income tax
 - c. an excise tax on particular products
 - d. a tax on business profits

Copies of the full questionnaire are available from Science Research Associates, 259 East Erie Street, Chicago, Illinois, which handles distribution of the questionnaire. SRA also has complete information on norm groups, plus an item analysis of all questions included.

II. What Economics Do High School Students and Teachers Know?

How much economics do high school students and their teachers know? And how does their economic understanding compare with that of other groups, such as college students?

Table 1 provides a summary picture, using test scores on the very elementary "Test of Economic Understanding" described above. While this test provides only a very limited measure of economic understand-

TABLE 1—LEVEL OF ECONOMIC UNDERSTANDING OF HIGH SCHOOL SENIORS AND OTHERS^a

	Test Scores	Standard Deviation	Significance of Difference from Preceding Group ^b
High school seniors: ^c			
No econ. courses (n=4601)	24.2	6.67	
One econ. course (n=1834)	29.7	8.19	.001
High school social studies teachers: ^d			
No college econ. courses (n=717)	32.0	8.48	
1-2 college econ. courses (n=1859)	32.8	8.42	N.S.
3-4 college econ. courses (n=1132)	35.5	8.94	.001
5+ college econ. courses (n=931)	37.2	8.48	.001
Watched "The American Economy" 3 or more times a week (n=110)	41.2	8.40	.001
College sophomores after econ. course: ^e (n=167)	40.7	4.20	
Industrial employees and managers: ^f			
Foremen and 1st line supervisors (n=319)	34.2	8.61	
Middle management (n=313)	36.3	8.13	.001
Staff: engineers, accountants, etc. (n=96)	36.6	7.11	N.S.
Top management (n=9)	42.7	8.4	.001

^a All test scores are on the standardized 50-item, objective "Test of Economic Understanding," except test scores for teachers are converted from a special 25-item form of the test.

^b Differences significant at .001 level where indicated. Differences marked N.S. were not significant at the .05 level.

^c Data from Science Research Associates on total national "norm group" for the "Test of Economic Understanding." Widely representative sample of about 6,500 students.

^d Data from NORC national stratified cluster sample of high school social studies teachers, described in text. Score adjusted to statistically comparable 50-item basis from 25-item test used by NORC.

^e Data from studies of Carnegie Tech and University of Nebraska students, reported by Phillip Saunders [8] and by C. R. McConnell and J. R. Felton [6], respectively. Standard deviation is for Carnegie Tech students only.

^f Data on special group of "industrial employees and managers" in 14 large, national companies tested by SRA for comparison purposes. Not necessarily typical of all companies. Data on top management shown merely for benchmark purposes.

ing, and while a more discriminating regression analysis controlling for such related variables as previous courses in economics is presented in Section VI, the basic scores are interesting, even striking for some categories.

The top two lines show the mean scores of some 6,500 high school seniors, divided between those who had taken no course in economics and those who had completed a one-semester course. The range of scores was from 8 to 48 on the 50-item test. While the students were not a scientifically drawn sample of all high school seniors, they were chosen to form a representative national "norm group" for users of the "Test of Economic Understanding." They comprise a roughly representative cross section of high school seniors, from large and small schools and cities, different sections of the country, and varying income areas.

As Table 1 indicates, even without a course in economics, high school seniors did roughly twice as well as the 12.5 score one would expect by chance, and taking a separate course in economics added 5.5 to the mean score of 24.2. This difference, which represents over a 20 per cent improvement, is statistically significant far above the .001 level, but it must be remembered that other variables (for example the possibility that brighter students tend to take separate courses in economics and the fact that economics courses are primarily offered in the "better" schools) may account for some of the difference. Even allowing for such special factors, however, it seems clear that a high school course in economics significantly increases students' ability to answer questions like those included on the "test of Economic Understanding."⁸

The next five lines show the test scores of the large NORC sample of high school social studies teachers. About 30 per cent of these teachers teach significant amounts of "economics" in separate courses or in courses in "problems of democracy"; the others teach primarily history, civics, and other social studies courses.

The mean score for all teachers was 34.1, with a range from 6 to 50 on a 50-item test basis.⁹ But this composite hides interesting differences. For example, high school social studies teachers who have never taken a separate course in economics (about one-sixth of the total) achieved a mean score of slightly over 32 out of 50. This was significantly above the score of high school seniors with no economics

⁸ For the results of a promising experiment with "programmed" teaching for high school students, see R. Attiyeh and K. Lumsden [1].

⁹ Although actually a shorter 25-item form of the test was used, this form was item-analyzed and was found to give virtually identical results with the longer version when the 25-item scores were converted to the 50-item basis.

but less than three questions better than high school seniors after they had taken a single course in economics.

Teachers with one or two previous one-semester college courses in economics (about 40 per cent of all social studies teachers) scored slightly higher, but, strikingly, the difference was not statistically significant. More than two college economics courses raised the test scores in the following groups. But five or more college courses were required to raise the teachers' mean score by five points, the amount one recent high school course raised the high school students' scores from a lower base. Those teachers who regularly watched the national television course, "The American Economy," three times or more a week averaged 41.2, considerably higher than teachers who had had five or more previous economics courses. But most of the regular watchers had also had three or more previous college courses in economics. One must remember also that the national television course was fresher in their minds (completed about nine months before the test was given), and other variables may have significantly affected the test scores. Section VI reports the results of a multiple regression analysis to isolate the impact of the TV course on test scores, holding other presumably relevant factors constant.

The bottom sections of Table 1 show the scores of Carnegie Tech and University of Nebraska sophomores who had just completed a regular course in economics or the national TV course, as reported by McConnell-Felton [6] and Saunders [8]. The mean score of approximately 41 was virtually identical for students in both schools, and for those who took a regular lecture-discussion course or the national TV course. Last, data are shown for samples of industrial supervisors, staff workers, and managers. The sample, drawn from 14 large companies, is not necessarily representative of all industrial firms, but it appears to be a roughly representative cross section of larger, reasonably well-known companies. The 42.7 mean score for top managers rests on a number of responses far too small to be reliable, but it is included for its casual interest. The other scores rest on respectable samples. Again, other variables would need to be analyzed to ascertain the precise significance of the scores, but they do present a crude basis for comparison with the scores for other groups shown.

Beyond these summary measures, detailed analyses of student and teacher performance on individual questions are available. By and large, students without economics training in high school missed questions indiscriminately, with no clear pattern as between subject matter areas or factual versus concept versus problem and application questions. Questions on monetary policy and operations, international economics, and comparisons between the Soviet and the U.S. economic

system were heavily missed, although teachers report (see below) that they consider comparative systems the most important area in economics to teach. Students generally got the central notions of consumer sovereignty and the role of competition in our system, and did surprisingly well on simple supply, demand, and price questions even without formal economics training.

There was no clear pattern in the improvements induced by studying a course in economics in high school. Scores improved on all types of questions, but especially on factual questions involving general magnitudes, on the comparison between the U.S. and the Soviet economies, and somewhat more on micro questions than on macro. The questions still missed after a course in economics tended to be the "harder" ones by analytical standards. But this was not a clear pattern. Even after a course in economics, nearly the same proportion still missed a simple question asking who is hurt most by inflation (farmers, debtors, government bond holders, or businessmen), which was missed by 76 per cent of the students. The balance of payments and monetary policy continued to be enigmas to a large share of the students.*

High school teachers, like their students, did better on micro than on macro questions, and did better on factual than on analytical questions. Beyond these general observations, no clear pattern emerged from the item analysis. As with the students, teachers having formal courses in economics showed general improvement, especially on "concept" and analytical questions. But the pattern of improvement was not a clear one.

III. *Current Courses and Teaching Practices*

A. *Courses and Textbooks*

Of the 12 million students in high school this year in the United States, somewhere between 10 and 20 per cent will take a separate course in economics before graduation. No completely reliable data exist on the figure, but three independent estimates based on large samples all fall within this range.⁸ Put the other way round, 80 to 90

* Tables showing a complete item analysis of the 100 questions on the two test forms are available on request.

⁸ A U.S. Office of Education total count for 1961-62 shows about 290,000 students registered in separate courses of economics during the year, of a total of about 2 million seniors. About 220,000 were in required courses, the balance in electives. But this questionnaire information leaves a substantial question as to just what is reported as a course in "economics." For example, "consumer economics" is probably included by many schools.

The Joint Council on Economic Education obtained substantially complete information on courses in economics in the 130 largest school systems in the United States and estimated on this sample (with adjustments for smaller systems based on uncertain evidence) that nearly 20 per cent of all seniors are taking a separate course in economics. Our estimate, making use of the NORC and other data, is more like 15 per cent.

per cent will graduate without having any formal instruction in economics per se.

In the 130 largest school systems in the country, approximately nine out of ten offer a separate course in economics of at least one semester, and over a quarter (including New York City) *require* a one-semester course in economics of all graduates. Nationwide, however, only about 40 per cent of all public high schools offer a separate course in economics, indicating that such courses are rare in the smaller schools outside major metropolitan areas. But there seems to be a clear trend in the direction of more required work in economics, in separate economics courses, and in other courses like problems of democracy and civics. Pennsylvania, for example, has recently mandated a requirement of 36 class hours of economics for every high school graduate in the state. About four-fifths of the 130 largest school systems reported curriculum revisions within the last three years increasing the amount of economics taught in high school, and the overwhelming majority of the teachers in the NORC study also reported plans to increase the emphasis on economics. Detailed information is presented below on the content of such courses and their success in raising the level of student understanding.

Beyond separate courses called "economics," apparently 15-20 per cent of all high school seniors take a one-semester or one-year course in "problems of democracy," in which there is usually at least one large separate unit on "economics" or some economic problem such as social security or natural resources. Few of these go beyond description of institutions and information on government legislation to deal with the problems faced. Nearly all states have a mandated course in "civics" or American government; many of these courses have units in "economics," mainly of a descriptive-institutional nature. On his way through high school, virtually every student must take a course in American history, and many schools claim that some economics is taught in American history courses. However, there is little evidence to back up this claim, unless one counts the fact that the student hears something about the growth of economic institutions such as business firms and the westward movement of the frontier. Most of the history books mention bimetallism, the Granger Acts, tariff legislation, and the like, as well as the development of agriculture and industry as part of the historical sweep which the student traverses, but there is little that the economist would recognize as economics. Lastly, many students in "business education" programs take a required course in "economics." Generally, this is heavily weighted with elementary personal finance, bookkeeping, office practice, and the like, although some broader courses are appearing.

Extensive field studies, textbook analyses, and reports of teachers generally confirm the picture of high school economics teaching. In a few schools, particularly in big cities and upper-class suburban areas, some very good high school economics is apparently being taught. But even in courses called "economics," the coverage is generally descriptive and nonanalytical. Much space is given to economic institutions, descriptions of natural resources, laws, governmental regulations, and the like. In the last few years, some texts have introduced some attention to problems of aggregate economics. Sometimes elementary economic concepts are introduced (for example, supply, demand, and price), but these are seldom used in application to the problems considered in the following chapters. At least to us, most of the texts seem taxonomical and probably dull for typical high school youngsters. As the AEA's Textbook Study Committee wrote: "Perhaps the most alarming characteristic of textbooks in all three courses [economics, social studies, and U.S. history] is the dominance of description over analysis in the treatment of those economic topics selected for discussion" [11, p. x].

B. Areas of Economics Taught in High Schools

High school teachers in the NORC survey described above were asked to select from a list of 11 subject matter areas in economics those which they considered to be "very important" to teach and those to which they actually devote at least six classroom sessions each year. Table 2 shows the results separately for those teaching courses in economics or problems of democracy (about 30 per cent of the total sample) and for all others teaching some economic topics, usually in civics, history, or general social studies courses. In reading Table 2 it should be remembered that, since most high school students do not take a course in economics, the attitudes and practices of teachers outside economics and problems of democracy courses may be more important on these issues than those of instructors teaching separate courses in economics.

Comparative economic systems (capitalism, communism, etc.) is considered far and away the most important area of economics to teach by both groups. About 90 per cent of all teachers considered it "very important," and about half report spending six or more periods on the subject.¹⁰ On the other hand, "the role of markets, prices, and

¹⁰ These high figures for comparative systems may reflect partly a tendency to put into this category ("capitalism, socialism, and communism") general treatments of the U.S. economy ("capitalism") which don't fit into any of the other categories, even when little attention is paid to other systems. However, a substantial portion of the states now require a course, or some minimum number of days, on comparative political, economic, and social systems, often with specific mention of communism and the USSR.

THE AMERICAN ECONOMIC REVIEW

TABLE 2—AREAS OF ECONOMICS EMPHASIZED BY HIGH SCHOOL TEACHERS*

Area	Rated "Very Important" by Teachers in:		Plan to Devote at Least Six Classes to:	
	(per cent)			
	Econ. & P.O.D.**	All Others	Econ. & P.O.D.**	All Others
1. Comparative systems (capitalism, communism, etc.)	91	87	53	47
2. Government finance, taxes, etc.	75	68	37	26
3. Labor unions, distribution of incomes	75	67	44	30
4. Development of economic institutions	70	64	41	31
5. Role of markets, prices, and profits	70	55	38	20
6. Booms, depressions, inflation, etc.	67	69	37	24
7. Government regulation of business	67	65	21	20
8. Consumer economics, personal finance	58	58	37	24
9. Money and banking	55	49	36	20
10. International economic problems	46	45	33	32
11. Underdeveloped economies	41	42	29	33

* Data from NORC. Table includes only teachers who reported that they teach something about economics, about 80 per cent of the total sample.

** Problems of Democracy.

profits" and "booms, depressions, and inflation" ranked well down the list among most social studies teachers, but, encouragingly, higher among those teaching economics per se. Interestingly, the distribution of income (labor unions, wages, social security, etc.) rated high among both groups, in contrast to what appears to be a tendency in the profession at the university level to play down this area in elementary courses. University economists will be interested in the place given the topic, "consumer economics, personal finance, etc." Although it ranked well down the list for both groups, it was still considered "very important" by more teachers than were money, international economics, and the underdeveloped economies.

C. Teaching Approaches and Attitudes

All teachers who cover any economics in their courses were presented three concrete topics (labor unions, the farm problem, and booms and depressions) and alternative ways of approaching each. They were asked to indicate which of three approaches they predominately use in teaching about each. One alternative stressed the historical approach, another the descriptive-institutional, and the third the development of economic concepts and their use in analysis of the situation. After this question was answered, teachers were asked specifically which of the three approaches they generally use the most.

BACH AND SAUNDERS: ECONOMIC EDUCATION

TABLE 3—TEACHING APPROACHES*

Approach	Economics and P.O.D. Teachers	All Other Teachers
Descriptive-institutional	6.3	6.3
Concept-analytical	5.8	5.4
Historical	4.8	5.2

* Based on NORC study. Each index is based on responses of that group on the teaching approach to each of the three problems presented, where 10 would represent use of that approach by all teachers on all three issues.

Not surprisingly, most teachers reported some use of all three approaches. Table 3 shows the relative stress placed on each of the three by teachers of economics and P.O.D., and by all others who teach anything about economics in their courses. The index number reported for each approach represents a weighted average of the number of first, second, and third choices each approach received relative to the others, where 10 would represent use of that approach by all teachers on all three topics.

As might be expected, the descriptive-institutional approach leads. More surprising is how closely the concept-analytical approach followed. But a warning is in order. All questionnaires were completely anonymous, and the questions on the treatment of labor unions, the farm problem, and economic growth and fluctuations carefully avoided such colored terms as "analytical" and "descriptive," so there should be little bias. But use of the "concept-analytical" approach in most cases implies only development of the simplest of economic concepts and their use in only the most elementary way. Other evidence, e.g., the AEA's textbook study quoted above, suggests little use of what economists would call "analytical" approaches to economic issues, no matter what the teachers replied on this part of the NORC study.

IV. The Teachers

Most of the economics taught in the high schools is offered in courses in "economics" or "problems of democracy." Table 4 presents information on the teachers of these courses and compares them with all other social studies teachers.

The table indicates that nearly all economics and P.O.D. teachers have had at least one college course in economics and that 58 per cent have had three or more.¹¹ A quarter of all economics and P.O.D. teachers have had five or more courses in economics, and 4 per cent

¹¹ It will be remembered from Table 1 that there was no significant difference on economics test scores for teachers with zero or one-two college courses in economics.

THE AMERICAN ECONOMIC REVIEW

TABLE 4—HIGH SCHOOL TEACHERS OF ECONOMICS AND PROBLEMS OF DEMOCRACY^a

	Economics and P.O.D. Teachers	All Other Social Studies Teachers
	(per cent)	
1. Number of college economics courses:		
None	10	18
1-2	32	43
3-4	32	21
5 or more	26	18
2. Major in: ^b		
Economics	6	3
Other social science (inc. history)	84	79
Other	38	37
3. Rank in college class:		
Top 10 per cent	26	18
Top quarter, not top 10 per cent	33	37
Second quarter	29	33
Bottom half	12	12
4. When took last economics course:		
Last year	11	5
2-5 years ago	32	30
5-10 years ago	22	30
More than 10 years ago	35	35
5. Score on "Test of Economic Understanding" ^c	36.9	32.9

^a Data from NORC survey.

^b Totals more than 100 per cent because some teachers reported more than one major, including both undergraduate and graduate levels.

^c Mean score on SRA test. For comparative data on other groups, see Table 1.

were majors in economics. Other social studies teachers have had considerably less economics.

High school social studies teachers as a group came from the top half of their college classes, only 12 per cent from the bottom half (though about two-thirds of the teachers obtained their grades through majors in "education"). Economics and P.O.D. teachers stood somewhat higher in their college classes than did other social studies teachers—26 per cent were in the top 10 per cent of their college classes. And over two-fifths of them were reasonably up to date, if recency of completing the last course in economics is a measure. Conversely, as indicated by item 4, nearly 60 per cent have not had a course in economics in the last five years, and a third not in the last ten years. The last line of the table shows that economics and P.O.D. teachers as a group did significantly better on the "Test of Economic Understanding" than did all other social studies teachers, as could be expected.

However, this comparison includes many other variables which influence test scores; a more thorough analysis, using a multiple regression technique, is provided in Section VI.

The NORC study also provides a large amount of additional information on high school social studies teachers. For example, 80 per cent of all such teachers were men. About one-third were under 30 years of age, another third between 30 and 44, and the rest older; the median age was 33.5. Interestingly, the proportion of these reporting no course in college economics, or only one to two such courses, was appreciably higher in the youngest age group than in the middle age groups. This suggests that the portion of potential social studies teachers taking courses in economics is lower now than it has been in the past, although it also reflects the fact that some teachers take economics courses toward advanced degrees after they begin teaching. About 40 per cent of all social studies teachers have earned some degree beyond the Bachelor's, and about 40 per cent are currently working toward some academic degree. Eleven per cent of all social studies teachers reported their college major was physical education.

Apparently about one-tenth of all social studies teachers are hired new each year. About a third have been teaching less than five years, while 43 per cent have been teaching at least 10 years. About 70 per cent report that all of their teaching is in the area of the social studies, with history the dominant area. About two-thirds taught at least one course in history, while about 13 per cent reported teaching a separate course in economics or economic institutions.

The median annual income of all social studies teachers in 1962-63 was \$6,150. About 25 per cent reported incomes under \$5,000 and 21 per cent reported \$7,500 or over. Their family backgrounds, as measured by father's occupation, conformed closely to the composition of the general population, except that more social studies teachers came from professional and fewer from farm families than in the general population. About 20 per cent of all social studies teachers reported reading the *New York Times* regularly; 5 per cent added the *Wall Street Journal*. Two per cent admitted to reading the *American Economic Review* regularly, while about 10-15 per cent reported regular reading of *Social Education*, *Social Studies*, or similar publications.

V. Who Watched "The American Economy?"

We turn now to an evaluation of the success of "The American Economy," the nationwide television course sponsored by the AEA in 1962-63. It was the most widely watched educational television course in history. Its total audience, averaging over one million persons daily, was apparently about twice as large as the highest previous audience

for a comparable national TV course, which was the course on probability and statistics broadcast in the preceding year. The some 5,200 viewers enrolled for credit at participating colleges was also the largest on record. Over 45,000 TV study guides for "The American Economy" were sold, and the National Educational Television Center reports that the sound films made from the TV tapes are the most widely used of any educational TV series ever produced—3,346 rentals and 531 sales of films in the course had been made as of November 30, 1964.

About 20 per cent of the 65,000 high school social studies teachers in the country watched the program at one time or another in 1961-63 or 1963-64. The NORC national sample survey, however, indicates that only about 5 per cent watched the program at least once a week throughout the 1962-63 year. A separate survey conducted by the National Association of Secondary School Principals of five large states (California, Connecticut, Illinois, Minnesota, and New York) indicated that approximately 15 per cent of the social studies and business education teachers in those states were watching the series "on a regular basis." Thus, it is clear that a substantial proportion of all social studies and business education teachers watched at least some of "The American Economy," but it seems probable that not more than 5-10 per cent of them (perhaps 3,000-6,000) were serious, regular viewers.

Some 245 colleges and universities offering a credit course based on "The American Economy" reported 5,200 students signed up for credit as of March, 1963. A subsequent postcard survey by the authors of this article indicates that approximately 85 per cent of these people (some 4,400) successfully completed the course for college credit. Of those completing the course, slightly over 40 per cent (some 1,800) were reported as school teachers. The remainder were regular undergraduate students taking "The American Economy" as their introductory economics course or other persons taking the course for credit.

Other surveys by the National Association of Secondary School Principals (nationwide) and the Committee for Economic Development (New Jersey), together with the NORC data, indicate that certainly over 1,000, and perhaps as many as 1,500, high school social studies and business education teachers successfully completed "The American Economy" for college credit.

A. Teachers Who Watched

What do we know about the high school social studies teachers who watched "The American Economy" regularly? Table 5 summarizes the answer, and compares these watchers with high school social studies teachers who were not regular viewers.

Perhaps the most striking finding is that two-thirds of all the high

school social studies teachers who watched "The American Economy" regularly at least once a week had previously had three or more courses in economics. Sixty-two per cent reported at least one graduate-level course in economics, and 44 per cent reported two or more such graduate courses (sometimes Schools of Education give graduate education credit for elementary work in economics when supplemented by advanced work in teaching methods). Only 14 per cent reported no course in economics. On the other hand, only 4 per cent reported that economics was the major field of their last academic degree. Item 3 confirms the related fact that the watchers were mainly people who were actively concerned with economics; more than half of all regular viewers were currently teaching a course in economics or problems of democracy.

This finding that half or more of the regular viewers had already had a substantial amount of economics accords with previous experi-

TABLE 5—HIGH SCHOOL SOCIAL STUDIES TEACHERS—VIEWERS AND NONVIEWERS^a

	Regular Viewers ^b	All Other H.S.S.S. Teachers
	(per cent in each category)	
1. Previous economics training:		
No economics courses	14	16
3 or more economics courses	67	42
Economics major for last degree	4	2.5
2. Advanced degree (beyond A.B.)	61	37
3. Teach a course in economics or P.O.D.	53	13
4. Standing in college graduating class		
Upper 10 per cent	27	20
Upper 25 per cent	65	56
5. Degree of professional activity ^c	14.4	10.1
6. Median age	43.8	33.4
7. Median years teaching	14.3	8.3
8. Sex (per cent male)	73	81

^a Based on NORC sample. Per cent in each case shows percentage of all teachers specified by the column heading, except for items 5-8, which are actual numbers.

^b Those who reported watching the program regularly once or more weekly throughout the 1962-63 year.

^c Weighted index of four measures of professional activity, including (a) the number of professional organizations to which the teacher belonged; (b) the number of times he has held office in a professional organization; (c) the number of professional and academic meetings attended over the past year; and (d) the number of professional and technical periodicals which the teacher reads regularly. Performance on each measure was coded 0-6, and possible scores on the index run from 0 to 24.

ence with educational television and related mass media. Previous studies have found that the role of such educational media is more one of reinforcing and supporting existing attitudes and interests than in developing new ones. People select from their environment stimuli that are meaningful to them in terms of previous experiences. Furthermore, recent studies in the field of adult education and of audiences for educational television in other fields show that those who participate in such programs and watch educational TV are more likely to be those who start with higher educational levels [5, pp. 80 and 136], [7, p. 57].

Lines 4 and 5 of Table 5 suggest that regular watchers ranked somewhat higher than other high school social studies teachers in academic standing, and that, as might be expected, they were generally more active in professional activities. Regular viewers were older and more experienced than were other teachers, and women comprised a substantially higher proportion than of all social studies teachers.

VI. Effectiveness of *"The American Economy"*

A. Economic Understanding

How effective was *"The American Economy"* in adding to the economic understanding of its viewers? The NORC study asked all of the social studies teachers who viewed the program at all whether it "added a great deal," "added somewhat," "added a little," or "didn't add anything." About 40 per cent of regular viewers reported that the program added a great deal to their understanding, and another 45 per cent that it added somewhat. Conversely, only 15 per cent considered that the program added little or nothing. Somewhat surprisingly, regular watchers who had already had three or more courses in economics felt that the program added just as much as did those who had had fewer or no courses at all in economics.

Such general reactions, however, are suspect as evidence of the actual learning that occurred from watching. As part of the NORC study, therefore, each social studies teacher in the sample was given a shortened, 25-question version of the "Test of Economic Understanding" described above. Performance on this test may serve as a rough measure of the economic understanding of each teacher. Therefore a multiple regression analysis was run to isolate the relative importance of watching the TV course and of some eight other variables in explaining performance on the test.¹² The major variables used in the

¹² As was explained above, scores on the 25-item test were converted to a statistically identical 50-item basis by multiplying by 2, to maintain comparability with the other scores shown in Table 1.

analysis were: watching the TV course; taking it for credit; previous training in college economics; teacher's standing in college graduating class; whether or not respondent teaches a separate course in economics or P.O.D.; teacher's professional motivation; and the personal characteristics of age and sex. The regression equation was of the usual linear form, $Y = a + bX_1 + bX_2 + \dots + bX_{14} + u$, where u is a random disturbance term assumed to have the usual simplifying properties.

Table 6 presents the results of the regression analysis. To provide extra information several of the main independent variables were subdivided into additive subvariables. For example, regular watchers of "The American Economy" were divided into "one or more times a week" and "three or more times a week," where all of the second group is included in the first. Thus, the coefficient in column 1 of Table 6 for "watched one or more times a week" is to be interpreted in the usual fashion as the effect of this variable, holding all others constant. The coefficient for "watched 3 or more times a week" also shows the effect of this variable, holding all others constant; thus it shows the *marginal* effect of watching 3 or more times over 1-2 times a week. To obtain the full impact of watching 3 or more times a week, we must add the two coefficients (.64 + 7.24), which gives us 7.88. Since this marginal analysis is applied for all of the first four major variables, column 2 has been added to show the full (summed) effect of the final subvariable in each group.

Table 6 indicates that watching "The American Economy" regularly three or more times weekly was far and away the most important variable in raising teachers' performance on the test of economic understanding. Its coefficient of 7.88 was more than twice as large as that of about 3.6 for having taken five or more college courses in economics or for graduating in the top 10 per cent of one's college class (the best proxy we had for intelligence, though one which also includes other factors such as motivation). It was much more powerful than taking one or more, or even three or more, college courses in economics. No other variable approached these in positive explanatory power.

The R^2 for the multiple regression is .152. For economists used to working with time series, this will seem extremely low. However, it is roughly in line with the R^2 's obtained in many other cross-section studies, for example of consumption behavior, and the F -test shows it is significant beyond the .001 level. The low R^2 may occur because major variables have been completely omitted—although it is hard to see what they might be. More likely, it is because of the large amount of random noise in such a large sample, and because the proxies used

THE AMERICAN ECONOMIC REVIEW

TABLE 6—MULTIPLE REGRESSION ANALYSIS OF RELATIVE INFLUENCE OF SELECTED VARIABLES ON TEACHERS' TEST SCORES^a
 $R^2 = .152$

Explanatory Variable	Marginal Subgroup Coefficients ^b	Coefficients ^b
Watched "The American Economy"		
X_1 = one or more times a week	.64	
X_2 = three or more times a week	7.24**	7.88
Took "The American Economy" for credit ^c		
X_3 = in-service credit or college credit	-8.90**	
X_4 = college credit only	3.10	-5.80
College economics training		
X_5 = one or more courses	.26	
X_6 = three or more courses	.88*	
X_7 = five or more courses	2.44**	3.58
College class standing		
X_8 = top 25 per cent	.12	
X_9 = top 10 per cent	3.48**	3.60
Teach high school economics or P.O.D.		2.36**
X_{10} = teach such a course		
Professional motivation		
X_{11} = activity in professional organization		-.12**
X_{12} = has or is working for advanced degree		1.16**
Personal characteristics		
X_{13} = sex (male)		2.22**
X_{14} = age		-.04**

^a Based on NORC data for 3,966 teachers; some responses could not be used because of incomplete data. See text for description of test of economic understanding used.

^b Using one-tailed *t*-test, * = significant at .01; ** = significant at .001. For difference between two columns, see text. All variables are dichotomous (0-1) except for age and for professional motivation (a continuous variable described in Table 5.)

^c "In-service" credit is usually offered directly by high school system toward salary increases; amount of work varies widely. "College" credit is formal credit, usually involving some on-campus review in teaching sessions plus formal examinations.

for important explanatory variables (for example, for intelligence and motivation) are imperfect; we comment below especially on what we suspect are inadequate measures of motivation. Similarly, the equation weights any college economics course as the equivalent of any other. But the proxies all appear to be reasonably sensible, and at least to give a tentative answer as to the relative importance of watching "The American Economy" as against other obvious possibilities for explaining performance on the test used.

The large negative coefficient for taking the TV course "for credit"

is one major surprise of the data. One might suppose that taking the course for credit would reflect high motivation, in addition to providing some additional education through supplementary on-campus or in-service classes. But taking the TV course for "in-service credit" subtracted nearly 9 points on the test score, holding all the other variables constant. Those taking the course for college credit did 3.10 points better than "in-service" credit-takers, but this difference was not statistically significant and, as indicated in column 2 of Table 6, it still left them with a coefficient of -5.80 .

This finding strongly suggests that the teachers taking "The American Economy" for credit differed markedly from other teachers on some other variables not included in our regression; otherwise one must conclude that the supplementary college teaching and examinations usually required for students taking the course for college credit actually confused the teachers and detracted from what they would have learned from just watching the TV course.

To investigate this unexpected result further, we examined directly all the cases of teachers taking the course for credit, only 42 individuals in the sample, of whom 23 were college-credit viewers and the others in-service credit viewers. This examination throws grave doubt on the superior motivation hypothesis for these watchers; about 25 per cent of the credit-takers reported watching less than three times a week. Indeed, it seems probable that many credit-takers may have wanted the credit more than the knowledge and did as little as possible to get by. A strong bimodal distribution of test scores among credit-takers supports this hypothesis. Several of the low scores came from the irregular watchers just noted, and the weighting system implied by the sampling procedure happened to give substantial weight to a few of these low test scores in producing the results shown. Clearly the "professional motivation" variables (X_{11} and X_{12}) are weak surrogates for the real motivational differences that may have existed. And one might reasonably hypothesize that a strong unmeasured motivation factor for regular viewers as a group is a major explanatory variable not picked up by the present regression equation.

There is additional evidence on the impact of the TV course, as measured by the same test. McConnell-Felton [6] and Saunders [8], using independent data in Pennsylvania and Nebraska, recently found that regular college students taking a regular year-long on-campus sophomore college course in economics scored about the same on the "Test of Economic Understanding" as did students taking "The American Economy" for college credit at the same institutions. McConnell-Felton reported, however, that on a more sophisticated test,

requiring more advanced technical tools from economics, regular college students taking a typical sophomore course outperformed the TV students by a significant margin. In further work, not yet published, Saunders reports a clear marginal improvement in the test performance of teachers who have taken the TV course and have supplemented this by on-campus classroom work with regular university instructors at some institutions. These results suggest the importance of high-quality instruction if classroom experience is to have any value in adding to a good TV presentation.¹³

B. *Residue from Previous College Economics Courses*

The coefficients for X_3 and X_4 in the regression equation indicate that one or two previous courses in college economics made no significant contribution to performance on the test of economic understanding, while three or four courses added very little to test performance. Five or more college courses added 3.58 to the test score, but even this was only about half the contribution made by watching "The American Economy" regularly during 1962-63.

This result may be interpreted by some as a devastating commentary on the effectiveness of our elementary economics courses or indeed on college economics as a whole. At least, it deserves further consideration.

First, on the average, teachers have been out of college about eight years, and the absence of any residue from college economics courses may merely reflect the well-known phenomenon of forgetting. Indeed, psychologists have shown that retention of most learned material is very short unless the material is deemed relevant by the learner and is used or otherwise reinforced periodically.¹⁴ Perhaps students retain virtually nothing from their other college courses either. But this is slight consolation if we measure the value of our courses by what lasts after the student escapes the final exam. The fact that we *may* do as well as other disciplines is hardly a happy defense. Nor does the fact that economics is not "used" or "reinforced" provide much of a rationalization, since everyone in this sample is an active teacher in the social studies, of which economics is surely one important component. Indeed, over 20 per cent are currently teaching a course in economics or problems of democracy. As is indicated by the coefficient for X_{10} , this current involvement with economics helps appreciably to raise

¹³ Preliminary findings from Saunders' further studies also show that this superior understanding of well-trained high school teachers is directly reflected in superior test performance of their high school students, as compared to control groups of students with other teachers.

¹⁴ See, e.g., Carl Hovland on "Learning," in *Handbook of Experimental Psychology* [14].

those teachers' test scores, but not enough to change the general picture significantly. It would have been remarkable had a college economics course taken (on the average) several years ago showed as much effect as the recent comparable TV course, and some allowance for this fact is required in assessing the relative success of the TV experiment; the McConnell-Felton and Saunders studies cited above provide direct evidence on this point. But this does not alter the basic finding of *no* significant residue from even a year of college economics.

Second, the absence of a residue from basic college courses may reflect the fact that the test used is a bad measure of economic understanding or that it measures the wrong things. Readers are invited to review the sample questions above and form their own judgments. Certainly the test is extremely elementary, since it was constructed to test basic, though nontechnical, aspects of economic understanding. It certainly will not discriminate effectively among people who know a good deal of economics—though a look back at Table 1 shows that mean scores for all the relevant groups are well below 100 per cent accuracy. Moreover, since it was designed to avoid rewarding mere acquaintance with technical textbook terms, this test may not show as strong an advantage for formal education types of information and understanding as would some other tests.

To obtain a further evaluation of the reasonableness of the "Test of Economic Understanding" for this purpose, the department chairmen of 30 leading universities and about 30 other leading economists especially interested in basic economics were asked to take the test and to write a brief impressionistic evaluation of it for the purpose for which it was designed. About half did so. Every reply stated that the test seemed at least "satisfactory" for the purpose indicated; most stated that it seemed "good," "highly appropriate," or "very good," although a number expressed reservations about individual questions. But of course the test may be a "good" one for its primary purpose and still not be satisfactory for evaluating desired lasting effects of our college courses.

Third, perhaps the minor carry-forward from college courses may reflect the fact that these teachers took unusually "poor" or "weak" undergraduate courses, and that students who took better courses would have performed significantly better on the test given. Since respondents' forms indicate where they studied as undergraduates, we have underway a supplementary analysis of this possibility. Preliminary data indicate that of about 4,200 teachers for whom information is available, about 1,300 (somewhat less than one-third) attended "teachers colleges." About 900 attended a group of 120 top "prestige" or very well-known universities and liberal arts colleges. About 2,000

(just under half the total) attended other colleges and universities. We also have preliminary information on how many majored in "education"; such majors accounted for about two-thirds of the total sample.

Economists who mistrust the educational standards of departments and schools of education may suspect that these data go far to explain the results reported above. Very preliminary analysis suggests that, by and large, noneducation majors did somewhat better on the test than their education counterparts in schools of comparable stature, and that high school teachers from the better-known schools did substantially better than those from other schools. But this, of course, may merely mirror differences in basic student abilities, and careful study holding such other variables constant will be required to judge whether in fact different types and sizes of programs and institutions appear to achieve significantly different lasting effects from their economics courses. We hope to report to the profession separately on this analysis in the near future. But while this further analysis may show significant differences in the success stories for different types of programs and institutions, the preliminary data suggest little reason to suppose it will change the basic picture presented here of generally low carry-forward from basic courses in college economics.

C. Plans for Change and Teaching Approaches

Teachers in the NORC sample also reported on plans to change the teaching time they will devote to different areas of economics. Nearly twice as many regular watchers of "The American Economy" (defined as those who watched three or more times a week) reported plans to increase the time spent on half or more of these areas next year as did nonwatchers. For example, 33 per cent of all regular watchers plan to increase the amount of time spent on six or more of the 11 areas of economics listed in Table 2, as compared to only 15 per cent of the nonwatchers. It is not, of course, permissible to attribute this difference solely to watching "The American Economy." Watchers may have been the ones who were inclined to put more time on economics in any case. However, the results are consistent with the hope that "The American Economy" would stimulate more attention to economics in high school social studies teaching.

Subject to the same reservation about causation, it is interesting that regular watchers show especially large increases in time spent on the core materials of macro- and microeconomics, compared to other teachers. For example, 39 per cent of all regular viewers are spending significantly more time on "economic stability and growth" than before the course, compared to only 22 per cent of other teachers. Simi-

larly, 37 per cent of the regular watchers now spend significantly more time on "the role of markets, prices, and profits," compared to 23 per cent of other teachers. Roughly comparable results were reported on "the development of modern economic institutions." In other areas of economics, the differences between the plans of watchers and non-watchers were much less marked, except that 33 per cent of the regular viewers reported plans to spend *less* time on consumer economics, compared to much smaller changes for other teachers. These results are consistent with the hope that "The American Economy" would develop more understanding of the central analytical core of economics and the way it can be used in thinking about economic problems.

Similarly, regular viewers reported a much greater emphasis on "analytic" (as contrasted to "descriptive" and "historical") teaching approaches than did occasional or nonviewers. On the three situations given (see III C), 24 per cent of regular viewers chose the analytical approach in teaching on all three, as compared to only 9 per cent of the other teachers. And 43 per cent of regular viewers chose an analytical approach to two of the three situations, as compared to only 27 per cent of the occasional or nonviewers. Again, the results are consistent with the hopes of "The American Economy" to stimulate a more analytical approach to economic issues, although they certainly cannot be attributed solely to that course.

This evaluation of the effectiveness of "The American Economy" does not include a large amount of *ad hoc* evidence reported by teachers, school administrators, and others interested in economic education from around the United States. These reports, almost without exception, agree that "The American Economy" was a widely watched, popular, effective TV program and course in economics. An abbreviated 60-film condensation of the course has been widely used in "in-service" teacher training programs and is now being used as the foundation for teacher development in 50 major school systems, under a new program by the Joint Council on Economic Education to help improve economic teaching in the schools. Experiments are also under way for the direct use of the films in high school courses. College teachers are using selected films widely, and industrial firms are using parts or all of the course for in-company development of middle- and lower-management people.

VII. *Some Implications*

In conclusion, we suggest the following as some implications of the findings.

1. If we want most of our future citizens to have any formal training in economics, it must be given in the high schools, barring an enor-

mous change in the national educational pattern beyond even the large increases in college enrollment currently expected. Thus, unless the profession wishes to wash its hands of responsibility for economic understanding of the citizenry, it must take a strong, active interest in the teaching of economics in the high schools.

2. It is possible to teach a substantial amount of economic understanding to average students in the high schools. Even with present inadequate high school courses in economics, students taking such courses showed large improvements in average test scores on the simple "Test of Economic Understanding." Still unpublished experiments in particular localities confirm this possibility, and indicate that with *well-trained* high school economics teachers or with effective "programmed learning" the improvement can be much more dramatic than shown in Table 1.

3. Better-trained high school teachers are critical in improving economic understanding provided by the schools. The small test margin of the mass of social studies teachers over average high school students who have had merely a weak one-semester course in economics is dramatic evidence on this point. Superintendents and other school administrators repeatedly stress the importance of improving the basic economic understanding of their social studies teachers if real improvement is to be made in their teaching. As indicated above, recent experiments confirm this strongly. Intensive work with competent, interested, and understanding university economists, followed up by in-service help, can dramatically improve the understanding of average high school teachers, their ability to teach effectively, and the performance of their students. However, merely taking more courses in economics or going through weakly taught summer institutes or in-service programs apparently does little good for high school teachers; quality of instruction and teaching materials appear to be crucial.

4. Unless the results reported above are grossly misleading, it is clear that present (or previous) college courses in economics don't do an effective job of preparing school teachers to teach economics, even recognizing the reservations indicated above. Whatever our students do on the final exam, the several-years-after test shows little residue, even for high school teachers for whom economic issues provide a part of their day-to-day teaching responsibilities. These findings emphasize again the well-known psychological principle that "learning" unsupported by motivation and reinforcement through repeated use or other means has a very short half-life. If our college courses don't develop student interest in economics for the years to come and if the analysis we teach isn't usable *and used* by students on their own after college,

there is little reason to expect much to last, however elegant the analysis or important the descriptive material in the course.

5. Since the average age of high school social studies teachers is only 33, and since about one-third of all teachers have been teaching less than five years and nearly two-thirds less than 10 years, improvement in the economic training provided in the colleges and universities could have a rapid impact on teaching in the high schools.

6. Improved textbooks and other teaching materials are critically needed as a foundation for improved teaching of economics in the schools. This includes not only materials for special courses in economics, but at least equally better materials for courses in problems of democracy, civics, American history, and the like. It is essential to remember that the great bulk of students get their exposure to economic issues in such courses. The economic preparation of the teachers in such courses is particularly weak, and such teachers badly need the best teaching materials.

7. If we want to get more economic analysis and points of view into history, problems of democracy, and civics courses in the schools, growing experience suggests that such teaching materials must be fitted into the patterns of those courses. For example, simply preparing booklets on economic analysis or description of economic institutions to be included in courses in American history or civics is unlikely to have much influence. Conversely, carefully developed materials which fit into the pattern of the American history course and develop important economic concepts and ways of using those concepts within the flow of the history course have been found valuable by history teachers.

8. Over all, there is little likelihood that economic understanding in the high schools will improve greatly unless school administrators and teachers get more sympathetic and active aid from professional economists than they have had to date.

REFERENCES

1. R. ATTIEH AND K. LUMSDEN, "The Effectiveness of Programmed Learning in Economics," *Am. Econ. Rev. Proc.*, May 1965, 55, 549-55.
2. ANN F. BRUNSWICK, *Economics on TV: An Evaluation of "The American Economy."* National Opinion Research Center Report No. 100, December 1964.
3. J. R. COLEMAN, "Economic Literacy: What Role for Television?" *Am. Econ. Rev. Proc.*, May 1963, 53, 645-52.
4. ——— AND K. O. ALEXANDER, *The American Economy: College of the Air Television Study Guide*. New York, 1962.
5. J. W. C. JOHNSTONE, *Volunteers for Learning*. National Opinion Research Center Report 89, 1963.

THE AMERICAN ECONOMIC REVIEW

6. C. R. McCONNELL AND J. R. FELTON, "A Controlled Evaluation of 'The American Economy,'" *Am. Econ. Rev.*, June 1964, 54, 403-7.
7. I. DE S. POOL AND B. ADLER, "Educational Television: Is Anyone Watching?" *Jour. of Social Issues*, April 1962, 18, 50-61.
8. PHILLIP SAUNDERS, "The Effectiveness of 'The American Economy' in Training Secondary School Teachers," *Am. Econ. Rev.*, June 1964, 54, 396-403.
9. *American Economic Review*, June 1964, 54, 565.
10. "College of the Air's 'The American Economy,'" *Am. Econ. Rev.*, September 1962, 52, 940-45.
11. "Economics in the Schools," *Am. Econ. Rev. Supplement*, March 1963, 53, Part 2, 1.
12. *Economic Education in the Schools: Report of the National Task Force on Economic Education*. New York, Committee for Economic Development, 1961.
13. *Economic Literacy for Americans: A Program for Schools and for Citizens*. New York, Committee for Economic Development, 1962.
14. *Handbook of Experiment Psychology*. New York 1951.

The Lasting Effects of Introductory Economics Courses

How much economics do students learn in "typical" two-semester sophomore introductory economics courses? How much of this learning lasts five years after these students graduate from college (some seven years after a sophomore economics course would have been taken)? How do students "feel" about the interest and difficulty of these courses? Do these feelings change as time goes on? Increasing numbers of calls for "accountability" in all the disciplines of higher education aside, these are important questions for economics—a profession with a particular concern for problems of efficiency and resource allocation.

On most college campuses more students are enrolled in introductory economics courses than in all other undergraduate economics courses combined, and academic members of the profession collectively spend more time teaching introductory economics than any other single course. Yet, the one most widely publicized pronouncement on the subject continues to be the now-famous "Stigler hypothesis" which suggests that, if an essay test on current economic problems were administered to college seniors (or persons five years out of college), there would be no difference in the performance of students with a "conventional" one-year course and those who had never had a course in economics (Stigler 1963).¹

If substantiated, the results predicted by Stigler would constitute a serious indictment of the pedagogical effectiveness of the profession. However, the kind of essay examination he proposed is not practical with the type of large nationwide sample that would be necessary to adequately test his hypothesis in its original form. Early empirical studies by Bach and Saunders (1965, 1966), using a multiple-choice examination, lent support to Stigler's hypothesis. However, these studies were limited to a nationwide sample of high school social studies teachers, and no data were available on the teachers' scholastic ability or the grades they received in their college economics courses. Also, the test used in the Bach-Saunders studies was the very elementary *Test of Economic Understanding* (TEU 1964), an instrument designed primarily for use with high school students.

This paper reports the results of a study designed to provide a more adequate, albeit still indirect, test of the Stigler hypothesis and to obtain additional information on college students' attitudes about economics and on their reading habits. The regression analyses reported below indicate that

Phillip Saunders is professor of economics at Indiana University. He acknowledges helpful suggestions and comments on an earlier draft of this paper from G. L. Bach, Rendigs Fels, W. Lee Hansen, and John J. Siegfried.

WINTER 1980

Stigler's original prediction may be unduly pessimistic. Introductory economics courses did have a lasting impact on students' performance as measured by a college-level multiple-choice test constructed especially for this study, but the lasting effects of introductory courses on test performance do appear to diminish over the years.

The study design and the survey instruments used are described in Section I. Test performance results are presented in Section II. Section III discusses some inconclusive results using additional variables for different types of introductory economics courses, and Section IV reports students' ratings of the introductory economics courses they took. Section V summarizes the major conclusions and discusses some implications of the study for the teaching of introductory economics.

I. Study Design and Test Instrument

This study compares the performance of students who have taken a "typical" two-semester college course in introductory economics with the performance of similar students who have not taken such a course. The comparisons were made at three different times: (1) immediately after an introductory course in economics, when most students are sophomores; (2) two years after an introductory course in economics, when most students are seniors; (3) five years after students have graduated from college—some seven years after a sophomore economics course would have been taken.

The data were collected during the 1969-70 and 1970-71 academic years from students and alumni at twenty-five carefully selected colleges and universities throughout the United States by means of a specially designed set of questionnaires. Part I of each questionnaire consisted of a series of questions concerning the respondents' sex, class standing or occupation, major course of study, whether they had taken introductory economics, their interest in economics as a subject, how important they thought economics was, whether they thought a course in economics should be required for college graduation, and a checklist of items designed to reveal their current reading habits. If the respondents indicated they had taken or were taking an introductory economics course, they were also asked to indicate whether it was required or an elective, and they were asked to rate the course, comparing it to other college courses they had taken, with respect to difficulty of subject matter, interest of subject matter, quality of textbook, quality of instruction, and time actually spent on the course.

The information obtained from Part I of the questionnaires was supplemented by data from school records on the respondents' SAT scores, grades actually received in all undergraduate economics courses taken, method of instruction (large sections, small sections, use of graduate student instructors, etc.), and textbooks used in the various introductory courses involved. Information was also obtained on the "intellectualism" (explained below) of the students at the schools included in this study.

Part II of each questionnaire was an especially devised version of the *Test of Understanding in College Economics* (hybrid TUCE). The test consisted of 33 four-option, multiple-choice questions selected from the four forms of the original *Test of Understanding in College Economics* (TUCE). The selected questions were designed to cover specified topics in micro- and macroeconomics using three different types of questions designated as recognition and

understanding (RU), simple application (SA), and complex application (CA).² Eleven questions of each type were included. A deliberate attempt was made to omit questions that seemed to rely on specific technical details, since it did not seem reasonable to expect alumni who had taken an introductory course several years previously to perform well on purely technical questions. The hybrid TUCE sought to measure economic *understanding*, not memorization. The RU questions cannot necessarily be answered by rote memory, and the SA and CA questions require respondents to *use* economics to arrive at the correct answer. Most of the CA questions are prefaced by actual or hypothetical newspaper quotations such as one is likely to be called upon to interpret and understand in real life outside the classroom.³

There were 1,220 sophomore respondents in this study, 955 senior respondents, and 1,257 alumni respondents. The sophomore respondents included 535 students (44 percent) with no college economics, and 685 (56 percent) with only a two-semester introductory course. The senior sample included 421 students (44 percent) with no college economics, 261 (27 percent) with only a two-semester introductory course, and 273 (29 percent) with up to nine one-semester college economics courses beyond the introductory level (the average for this group was 3.83 courses). The alumni sample included 435 respondents (35 percent) with no college economics, 464 (37 percent) with only a two-semester introductory course, and 358 (28 percent) with up to nine one-semester college economics courses beyond the introductory level (the average for this group was 3.50 courses).⁴

II. Test Performance of All Respondents

The results of the regressions run with the hybrid-TUCE score as the dependent variable are shown in tables 1 and 2. Since cross-sectional data were used, separate regressions were run for the sophomore, senior, and alumni samples. Two regressions were run with the alumni sample: a "short" one using exactly the same variables as for the sophomore and senior regressions; and a "long" one that included variables on marital status, occupation, and income, for which data were available from only the alumni respondents. Table 1 treats the respondents' experience in college economics courses only in terms of the number of economics courses taken. Table 2 also includes data for the grades received in each course. All course grades were converted to a scale on which A = 4, B = 3, C = 2, D = 1, and F = 0; and grades for the two semesters were averaged to get a single grade for the entire introductory course.

ECONOMICS EXPERIENCE

Introductory courses. As indicated in Table 1, holding other things constant, taking an introductory college economics course is significantly associated with a difference in total test scores of 6.18 points in the sophomore sample, 4.76 points in the senior sample, and either 3.23 or 3.24 points in the alumni sample, depending upon which alumni regression is used. Thus, the major finding of this study is that taking an introductory economics course does have both an immediate and a lasting effect on hybrid-TUCE scores.⁵ Furthermore, Table 2 indicates that, holding other things constant, each letter grade in introductory economics is significantly associated with a difference in total test score of 2.00 points in the sophomore sample, 1.69 in the senior sample, and either 1.11 or 1.09 in the alumni sample.

WINTER 1980

TABLE 1
Multiple Regression Results with Economics Experience by Courses
 (dependent variable: total Hybrid TUCE score with range of 33 to 4)

Variable and Code ^a	Sophomores (SO)		Seniors (SR)		Alumni (AL)			
	Regres. Coeff.	t Stat.	Regres. Coeff.	t Stat.	Regres. Coeff.	t Stat.	Regres. Coeff.	t Stat.
No. of respondents	1220		955		1257		1257	
R ²	548		595		550		538	
Stand. error of est.	4.126		4.111		3.734		3.777	
Economics experience:								
Course in intro. econ.	6.18	17.77**	4.76	8.75**	3.23	5.05**	3.24	5.06**
Intro. econ. req.	-.70	-1.85	-1.08	-2.61**	-.59	-2.02*	-.55	-1.88
No. of econ. courses beyond intro. (9-0)	-	-	.53	5.23**	.50	5.86**	.48	5.55**
Individual SAT total (1592-528)	1.17	13.84**	.83	12.17**	.95	12.45**	.91	11.83**
"Intellectualism" of student body (79-27)	.33	2.53*	.31	2.16*	.27	2.06*	.31	2.36*
Personal characteristics, attitudes, & reading habits:								
Male	.71	2.46*	1.04	3.17**	1.04	4.08**	.97	3.46**
Married	-	-	-	-	-	-	.42	1.64
Interest in econ. (5-1)	.98	5.64**	1.12	5.97**	.81	5.68**	.78	5.47**
Importance of econ. (5-1)	.04	.23	.25	1.29	-.19	-1.18	-.15	-.91
Make econ. req. (5-1)	-.07	-.59	-.43	-3.25**	-.24	-2.32*	-.25	-2.37*
Read bus. sect. daily paper (2-0)	.04	.20	.24	.98	.04	.26	.01	.06
Read <i>Wall Street Journal</i> (2-0)	-.04	-.16	-.36	-1.29	.12	.63	-.02	-.09
Read bus. sect. weekly news mag. (2-0)	.38	2.06*	.84	4.07**	.44	3.00**	.44	2.98**
Read <i>Business Week</i> (2-0)	-.16	-.51	-.89	-2.66**	-.10	-.50	-.11	-.56

Read <i>Fortune</i> (2-0)	21	.77	-.36	-1.31	44	2.28*	.38	1.99*
Read invest. adv. serv. (2-0)	-45	-1.36	.06	.19	-.06	-.36	-.07	-.37
Major course of study: ^b								
Nat. sci., math, or engin.	-33	-.85	.33	.76	58	2.14*	.29	1.03
Hum. or fine arts	-40	-.98	-.84	-1.88	-63	-2.07*	-.94	-2.92**
Economics	1.15	2.06*	.83	1.09	39	.53	-.13	-.19
Nonecon. bus. ad.	.53	1.14	.78	1.42	60	1.38	-.15	-.32
Education	-.79	-1.73	.11	.21	-1.17	-3.15**	-1.34	-3.17**
"Other"	-.44	-.83	-1.42	-2.32*	.17	.35	-.14	-.29
Occupation: ^c								
Econ.							2.17	3.50**
Econ. and bus. related							.79	1.82
Prim. or sec. teacher							-.47	-1.27
College or jr. coll. teaching, & grad. student							.61	2.00*
Housewife							-.03	-.08
Income: ^d								
\$25,000 and over							.52	.50
\$20,000-24,999							.35	.42
\$15,000-19,999							.64	1.35
\$10,000-14,999							-.12	-.45
Under \$4,000							-.07	-.21
No answer							1.63	2.15*

*Significant at .05 level.

**Significant at .01 level.

^aWhere not stated, code is 1-0, and 1 = yes.^bMajor in social science other than economics suppressed in intercept.^c"Other" occupations suppressed in intercept.^dIncome \$5,000-\$9,999 suppressed in intercept.

TABLE 2
Multiple Regression Results with Economics Experience by Grades in Courses
 (dependent variable: total Hybrid TUCE score with range of 33 to 4)

Variable and Code ^a	Sophomores (SO)		Seniors (SR)		Alumni (AL)			
	Regres. Coeff.	t Stat.	Regres. Coeff.	t Stat.	Regres. Coeff.	t Stat.	Regres. Coeff.	t Stat.
No. of respondents	1220		955		1257		1257	
R ²	.555		.619		.567		.557	
Stand. error of est.	4.092		3.990		3.663		3.687	
Economics experience:								
Grade in intro. econ. (4.0-0.0)	2.00	18.42**	1.69	11.43**	1.11	8.27**	1.09	8.11**
Intro. econ. req.	-.02	-.06	-.77	-2.00*	-.25	-.87	-.23	-.81
No. of econ. courses beyond intro. x avg.								
Grade in other econ. courses (36.0-0.0)	-	-	.18	5.22**	.16	5.73**	.15	5.26**
Individual SAT total (1592-528)	1.01	11.90**	.75	11.28**	.83	10.91**	.79	10.32**
"Intellectualism" of student body (79-27)	.41	3.22**	.41	2.95**	.38	2.90**	.41	3.19**
Personal characteristics, attitudes, & reading habits:								
Male	.96	3.38**	.97	3.06**	1.19	4.78**	1.06	3.89**
Married	-	-	-	-	-	-	.39	1.53
Interest in econ. (5-1)	.87	5.02**	.90	4.86**	.75	5.34**	.72	5.18**
Importance of econ. (5-1)	.06	.34	.26	1.41	-.24	-1.49	-.19	-1.23
Make econ. req. (5-1)	-.15	-1.22	-.42	-3.27**	-.24	-2.40**	-.25	-2.42*
Read bus. sect. daily paper (2-0)	.14	.61	.22	.94	.01	.09	-.02	-.13
Read <i>Wall Street Journal</i> (2-0)	.02	.07	-.42	-1.54	.15	.80	.03	.15
Read bus. sect. weekly news mag. (2-0)	.38	2.08*	.75	3.72**	.42	2.92**	.42	2.94**
Read <i>Business Week</i> (2-0)	-.17	-.55	-.82	-2.53*	.00	.01	-.01	-.07

Read <i>Fortune</i> (2-0)	.19	.72	-.25	-.92	.46	2.43*	.39	2.09*
Read invest. adv. serv. (2-0)	-.42	-1.29	.16	.52	.01	.07	-.02	-.14
Major course of study: ^a								
Nat. sci., math, or engin.	-.53	-1.37	.06	.14	.23	.85	.01	.03
Hum. or fine arts	-.59	-1.46	-.94	-2.19*	-.76	-2.54*	-.99	-3.15**
Economics	.46	.83	.47	.65	.29	.49**	-.05	-.07
Nonecon. bus. ad.	.49	1.08	1.02	1.94	.26	.62*	-.36	-.76
Education	-1.19	-2.62**	-.13	-.26	-1.36	-2.54**	-1.41	-3.41**
"Other"	-.77	-1.44	-1.52	-2.57*	.11	.23	-.15	-.31
Occupation: ^c								
Econ.							2.01	3.29**
Econ. and bus. related							.79	1.87
Prim. or sec. teacher							-.60	-1.64
College or jr. coll. teaching, & grad. student							.46	1.52
Housewife							-.22	-.56
Income: ^d								
\$25,000 and over							.40	.39
\$20,000-24,999							-.19	-.24
\$15,000-19,999							.45	.96
\$10,000-14,999							-.05	-.17
Under \$4,000							-.06	-.17
No answer							1.69	2.28*

*Significant at .05 level.

**Significant at .01 level.

*Where not stated, code is 1-0, and 1 = yes.

*Major in social science other than economics suppressed in intercept.

*Other occupations suppressed in intercept.

*Income \$5,000-\$9,999 suppressed in intercept.

The results in both tables indicate that the differences in test performance associated with introductory economics courses in the alumni sample are slightly over half as large as they are in the sophomore sample. Since cross-sectional rather than time series data are used, this *implies*, but does not demonstrate, that there is an overall "retention rate" of some 52 percent–55 percent (or an overall "decay rate" of 45 percent–48 percent) from the end of a sophomore course to a time some seven years later. This is equivalent to an *annual* rate of depreciation of economic knowledge of about 6 percent.

More disaggregated data not shown here indicate that between the alumni sample and the sophomore sample the implied "retention rate" is larger (and the implied "decay rate" is smaller) on the simple and complex application questions than on the recognition and understanding questions used on the hybrid TUCE.⁶

Courses beyond the introductory level. The data in tables 1 and 2 indicate that each economics course taken beyond the introductory level is associated with a difference in total test performance of 0.53 points in the senior sample and 0.50 or 0.48 in the alumni sample. A difference of one letter grade in each such course is associated with a difference in total test performance of 0.18 points in the senior sample and 0.16 or 0.15 in the alumni sample. As with the introductory course, this indicates a lasting impact for upper-level economics courses and course grades.

OTHER CONTROL VARIABLES. Data were collected on several other variables in addition to upper-level college economics courses, so that the influence, if any, of these variables could also be held constant in comparisons between respondents with and without introductory economics. Several of these variables reported in tables 1 and 2 were consistently and significantly associated with test performance in all three of our samples. Other control variables did not prove to be consistently or significantly associated with test performance in all three samples. Before discussing the various control variables, it should be noted that the R^2 s of the regressions shown in tables 1 and 2 are high compared to other cross-sectional studies done in economic education; but the variables in the regressions still explain only 54–62 percent of the variation in total test scores observed in the various samples.

Individual SAT scores. In all the regressions shown in tables 1 and 2, each 100-point difference in a respondent's SAT score is positively and significantly associated with a difference in total test performance. The size of the coefficient is roughly one point, and ranges from a high of 1.17 points for the sophomore sample in Table 1 to a low of 0.75 for the senior sample in Table 2.

"Intellectualism" of the student body. In addition to the respondent's SAT score, which was taken to be the best available proxy for an individual's innate intellectual "ability," we also obtained data on the intellectual "ability" of the entire student body of the sample schools to see whether, independent of a particular individual's ability, association with "bright" students or attendance at a "high-quality" school that attracts "bright" students and "good" faculty is a factor in economic understanding. This consideration was suggested in an earlier study by Attiyeh et al. (1971, p. 71) who found the average freshman entrance examination score in each school was highly significant and nearly half as important to an individual's performance on an economics test as his or her own SAT entrance examination score.

We used three measures of the character of each school's student body and conducted a sensitivity analysis to determine which of the three was the most significant for control purposes. One measure was the average SAT scores of entering freshmen. The other two were "selectivity" and "intellectualism" scales constructed by Astin (1965).⁸ The intellectualism scale proved to be the most useful for our purposes, but the overall results were not affected in any substantial way when either of the other two measures was substituted.

Tables 1 and 2 show that a ten-point difference in the intellectualism of a school's students is positively and significantly associated with a difference in total test performance in all the regressions shown. The size of the coefficients ranges from 0.41 to 0.27. Even for alumni, attendance at a school with an intellectualism rating of say 70 rather than 30 is associated with a difference in test performance of between 1.64 and 1.08 points, depending upon which other variables are introduced into the basic regression model. The nature of a school's student body thus appears to be a significant factor that should not be ignored in future studies attempting to explain economics test performance.

Gender. As has been true in many other studies (Siegfried 1979), males did significantly better than females on the economics test in all of the regressions shown in tables 1 and 2. The size of the coefficient for males ranged from 1.19 in the "short" alumni regression in Table 2 to 0.71 in the sophomore regression in Table 1.

General attitudes toward economics. All questionnaires contained the following three questions, with the responses coded as indicated by the numbers in parentheses, although the numbers themselves did not show on the questionnaire forms:

How would you rate your present *interest* in economics as a subject? (Check one.) Very High (5), High (4), Average (3), Low (2), Very Low (1).

How *important* do you think a general understanding of economics is in today's world? (Check one.) Very Important (5), Important (4), Fairly Important (3), Fairly Unimportant (2), Very Unimportant (1).

Do you feel that all students should be required to take a course in economics before they graduate from college? (Check one.) Strongly Agree (5), Agree (4), Undecided (3), Disagree (2), Strongly Disagree (1).

Each one-point difference on the scale for the first question was positively and significantly associated with a difference in total test performance in all of the regressions shown. The size of the coefficient ranged from 1.12 in the senior regression in Table 1 to 0.72 in the "long" alumni regression in Table 2. The responses to the second question were not significantly associated with test performance in any of the regressions shown. For the last question, a one-point difference on the scale was *negatively* associated with test performance in all the regressions shown. The coefficients, all of which were fairly small, were statistically significant in the senior and alumni regressions, but not in the sophomore regressions.

The positive association between interest in economics as a subject and test performance is not surprising, but it is more difficult to explain the weaker negative association between a belief that economics should be required and

performance on the hybrid TUCE. Perhaps an awareness of one's lack of economic understanding is associated with the belief that college students should be required to take economics courses.

Reading habits. All respondents were asked whether they read the economic, business, or financial sections of a daily newspaper; a weekly news magazine such as *Newsweek*, *Time*, or *U.S. News and World Report*; and the other items shown in tables 1 and 2. For each, a response of "read frequently" was coded 2; "read occasionally," 1; and "read never or hardly ever," 0. The only consistent, statistically significant association with test performance was for reading the economics section of a weekly news magazine.

Major course of study. With two exceptions, after adjusting for all the other variables in the regressions, a respondent's major course of study does not appear to have a strong influence on his or her test performance. Humanities and fine arts majors did have consistently negative coefficients, however, and five of these coefficients (including all four in the alumni regressions) were statistically significant. Five of the negative coefficients for education majors (again including all four in the alumni regression) were also significant statistically, and the size of these coefficients in the alumni regressions (-1.17 to -1.41) may help explain why the earlier Bach-Saunders studies, confined to social studies teachers, most of whom were presumably education majors, failed to find the significant lasting effect for introductory economics courses that appears in the present study.

Marital status, occupation, and income. Expanding the "short" alumni regression (third column in tables 1 and 2) by introducing the variables on marital status, occupation, and income increased the amount of variation explained only slightly (fourth column): R^2 increased 0.012 or 1.2 percentage points in Table 1 and 0.010 or one percentage point in Table 2. Readers can explore for themselves the influence of the individual items in each of these categories in tables 1 and 2. Of all of the professions examined separately, for example, only the economics profession did significantly better on the total hybrid TUCE than did the "other" professions suppressed in the intercept in both tables.

III. Test Performance of Respondents Taking Different Types of Economics Courses

Regressions similar to those reported above were run using information on class size, use of graduate student instructors, textbook used, and respondents' ratings of course difficulty, course interest, quality of text, quality of instruction, and time spent on introductory economics. None of these additional variables had a consistent, statistically significant association with test performance, however, and a major disappointment of this study is that we were not able to discern any particular types of economics courses, instruction, or textbooks that had a consistent, significantly different impact from any other type. There is apparently no "one best way" to lead all students to economic understanding.

IV. Student Ratings of Introductory Economics Courses

Another set of regressions was run using as dependent variables each of the five different scales on which the respondents, with introductory economics were asked to rate and compare these courses with other college courses they had taken. While there were again no consistent differences associated with these

TABLE 3
Simple Comparison of Mean Student Ratings of Introductory
Economics Courses in Different Samples*

(Question: Compared to other college courses you took, how would you rate your college introductory economics course on each of the following items? [5 = "much more" or "one of the very best," 4 = above average, 3 = average, 2 = below average, 1 = "much less" or "one of the very worst"])

Rating Items	Seniors				Alumni		
	Soph.	Intro.	Beyond	Total	Intro.	Beyond	Total
		Only	Intro.		Only	Intro.	
Difficulty of subject matter	3.27 (.83)	3.20 (.88)	3.29 (.80)	3.24 (.84)	3.04 (.94)	2.96 (.80)	3.00 (.88)
Interest of subject matter	3.28 (.97)	2.69 (.97)	3.54 (.94)	3.12 (1.04)	2.72 (1.01)	3.36 (1.01)	3.00 (1.06)
Quality of textbook	3.36 (.83)	3.20 (.93)	3.47 (.99)	3.34 (.97)	3.27 (1.00)	3.61 (.87)	3.42 (.96)
Quality of instruction	3.71 (.95)	3.18 (1.19)	3.67 (1.11)	3.44 (1.17)	2.98 (1.10)	3.32 (1.02)	3.13 (1.08)
Time you actually spent on the course	3.08 (.91)	2.85 (.94)	3.29 (.87)	3.08 (.93)	2.70 (.97)	2.96 (.81)	2.81 (.91)
No. of respondents	685	534	273	807	464	358	822

*Standard deviations are shown in parentheses below the means.

measures and the different "type-of-course" variables mentioned above, the data collected nevertheless provide some interesting information on students' opinion of the introductory economics courses offered during the 1960s (see Table 3).⁹

As indicated in Table 3, the respondents in all three samples who had taken a course in introductory economics were asked to compare this course with other college courses in terms of difficulty, interest, quality of text, quality of instruction, and time actually spent on the course. The five possible responses and their codes are listed in the table.

Most of the mean ratings in Table 3 are greater than the "average" value of 3.00, and the seniors and alumni who took additional courses beyond the introductory level consistently rated the course higher than those who took only the introductory course. Enough data are provided in the table to permit the calculation of tests of statistical significance for any particular comparisons in which the reader might be interested, but a few words of caution should be emphasized. Like the test scores discussed above, the student ratings should be interpreted on a cross-sectional rather than a time-series basis. Different people were rating the same courses at different times. Also, the framework of "other college courses" being used for comparison is larger for the senior and alumni respondents since it contains junior- and senior-year courses not yet taken by the respondents in the sophomore sample. The alumni also had to think back over a longer period of time than the seniors and sophomores. Finally, although our study indicated no major changes, the courses rated may themselves have changed somewhat over the years.

WINTER 1980

V. Concluding Comments and Implications

Teachers of introductory economics can take some encouragement from the results of this study. While there is little room for complacency, their efforts have not been completely in vain, and introductory economics courses do appear to have some impact that lasts beyond the final exam. A difference of over three points on the hybrid TUCE some seven years after a sophomore course has been completed is not only statistically significant, it also seems to be educationally important. This is over 50 percent of the difference associated with taking an introductory course immediately after the course is completed. This finding casts some doubt on the Stigler hypothesis quoted above. Although it is true that we did not use the type of essay questions he originally had in mind, Stigler himself was a member of the original TUCE committee, and thus had a hand in formulating the test questions used in this study. It is also true that the hybrid TUCE used in this study puts a much heavier emphasis on realistic application questions than any alternative testing instrument available.

While our findings cast doubt on the prediction of no difference in student performance, they do lend some support to a basic point Stigler was making in criticizing "the watered-down encyclopedia which constitutes the present course in beginning college economics" which, he argued, "does not teach the student how to think on economic questions" (p. 657). The implications in our results that suggest the retention rate is higher on SA and CA questions than on RU questions supports Stigler's statement that "an introductory-terminal course in economics makes its greatest contribution to the education of students if it concentrates upon a few subjects which are developed in sufficient detail and applied to a sufficient variety of actual economic problems to cause the student to absorb the basic logic of the approach" (p. 658). Our data also imply that introductory economics courses might well devote more explicit attention to increasing students' interest in economics as a subject and encouraging them to make a habit of reading the economics or business sections of weekly news magazines, since these factors appear to exert an independent influence on economic understanding. With economic events and reporting increasingly moving from the business page to the front page, we may have an opportunity to significantly extend the lasting effectiveness of our introductory economics courses.

FOOTNOTES

1. In Stigler's own words, the complete hypothesis was stated as follows: "I propose the following test: Select an adequate sample of seniors (I would prefer men five years out of college), equally divided between those who have never had a course in economics and those who have had a conventional one-year course. Give them an examination on current economic problems, not on textbook questions. I predict they will not differ in their performance. I shall illustrate below the kind of question that should be asked in this test" (p. 657). . . .

"Give the student a summary page or two of the arguments and evidence presented in the discussion (in Congress and the public press) of HR 5983 [regarding turkeys] and let him explain benefits and costs of the scheme—with the grading based, of course, on the coherence of his argument and relevance of his evidence, not on the conclusions reached" (p. 659).

2. The original TUCE was developed by a distinguished committee consisting of Rendigs Fels, Vanderbilt University, chairman; G. L. Bach, Stanford University;

William G. Bowen, Princeton University; R. A. Gordon, University of California (Berkeley); Bernard F. Haley, University of California (Santa Cruz); Paul A. Samuelson, Massachusetts Institute of Technology; George Stigler, University of Chicago; John M. Stalnaker, National Merit Scholarship Corporation, consultant; and Paul L. Dressel, Michigan State University, executive director. The four versions of the original TUCE are described and student performance data are discussed in Fels (1967), Psychological Corporation (1968), and Welsh and Fels (1969).

3. See Saunders and Welsh (1975). Copies of the test, questionnaires, and specific details of the sampling procedure used in this study are available in Saunders (1973) and can be obtained by writing to the author.
4. Students with only a one-semester introductory economics course, or students who took only one semester of a two-semester sequence, were omitted from the study, which was designed to test the effectiveness of the "typical" two-semester college introductory course.
5. The data in tables 1 and 2 also indicate that taking introductory economics on a required rather than an elective basis has a consistent negative association with total test performance, but only three of the eight coefficients were statistically significant at the .05 level.
6. A series of regressions identical to those shown in tables 1 and 2 was run with the scores on each of the sets of eleven RU, SA, and CA questions, rather than the total hybrid-TUCE scores, as the dependent variable. Of the total difference of 2.94 points associated with an introductory economics course between the sophomore regression and the long alumni regression ($6.18 - 3.24 = 2.94$) there was a difference of 1.25 points (42.5 percent of the total) on the 11 RU questions, a difference of 0.84 points (28.6 percent of the total) on the 11 SA questions, and a difference of 0.85 points (28.9 percent of the total) on the 11 CA questions.

Looked at in another way, on the 11 RU questions, the differences associated with an introductory course declined from 2.44 points in the sophomore regression to 1.19 points in the alumni regression (a change of 1.25 points or 51.2 percent of the sophomore difference); on the 11 SA questions the difference associated with an introductory course declined from 2.03 points in the sophomore regression to 1.19 points in the alumni regression (a change of 0.84 points or 41.4 percent of the difference in the sophomore regression); on the 11 CA questions, the difference associated with an introductory course declined from 1.71 points in the sophomore regression to 0.86 points in the alumni regression (a change of 0.85 points or 49.7 percent of the difference in the sophomore regression).

Similar results were obtained in regressions using course grades as the introductory economics variable.

7. Each SAT score has a maximum value of 1,600, the sum of the scores for verbal aptitude and for mathematical aptitude, each with a maximum value of 800.
8. Astin's "intellectualism" scale is expressed as a range of "T-scores" with a mean of 50 and a standard deviation of 10, and it is constructed such that "an entering student body with a high score would be expected to be high in academic aptitude (especially mathematical aptitude) and to have a high percentage of its students pursuing careers in science and planning to go on for Ph.D. degrees" (p. 54). The unweighted mean of the schools in the lasting effects study on Astin's intellectualism scale was 54.52, with a range of 79 to 27. This indicates that the schools represented in our study are slightly above the average of all U.S. colleges and universities in terms of the intellectualism of their student bodies; but our sample covers a broad range, and there is no reason to believe that it is not adequate or representative for our purposes.
9. A discussion of a partial preliminary version of these findings, along with a discussion of the methodological problems of comparing results from the different samples used in this study, appears in Saunders (1971) and a more extensive methodological discussion is contained in Saunders (1973).

REFERENCES

- Astin, A. W. *Who Goes Where to College*. Chicago, 1965.
- Attiyeh, R. E.; G. L. Bach; and K. G. Lumsden. "Factors Affecting Student Learning of Elementary Economics." In *Recent Research in Economic Education*, edited by K. G. Lumsden. Englewood Cliffs, N. J.: Prentice-Hall, 1971.
- Bach, G. L. and P. Saunders. "Economic Education: Aspirations and Achievements." *Amer. Econ. Rev.* 55 (June 1965):329-356.
- . "Lasting Effects of Economics Courses at Different Types of Institutions." *Amer. Econ. Rev.* 56 (June 1966):505-511.
- Fels, R. "A New 'Test of Understanding in College Economics'." *Amer. Econ. Rev., Proceedings* 57 (May 1967):660-666.
- Saunders, P. "The Lasting Effects of Elementary Economics Courses: Some Preliminary Results." *Amer. Econ. Rev., Proceedings* 61 (May 1971):242-248.
- . "The Lasting Effectiveness of Introductory Economics Courses: Final Report NSF Grant GY7208." Mimeo 1973.
- Saunders, P., and A. L. Welsh. "The Hybrid TUCE: Origin, Data, and Limitations." *J. Econ. Ed.* 7:1 (Fall 1975):13-19.
- Siegfried, J. "Male-Female Differences in Economic Education: A Survey." *J. Econ. Ed.* 10:2 (Spring 1979):1-11.
- Stigler, G. J. "Elementary Economic Education." *Amer. Econ. Rev., Proceedings* 53 (May 1963):653-659.
- TEU. *Test of Economic Understanding: Interpretive Manual and Discussion Guide*. Chicago: Science Research Associates, 1964.
- TUCE. *Test of Understanding in College Economics: Manual*. New York: The Psychological Corporation, 1968.
- Welsh, A. L., and R. Fels. "Performance on the New Test of Understanding in College Economics." *Amer. Econ. Rev., Proceedings* 59 (May 1969):224-229.

IS TEACHING THE BEST WAY TO LEARN? AN EVALUATION OF BENEFITS AND COSTS TO UNDERGRADUATE STUDENT PROCTORS IN ELEMENTARY ECONOMICS*

I. INTRODUCTION

The Joint Council on Economic Education (JCEE) has recently sponsored several projects to explore alternative approaches to teaching college introductory economics. The experimental course at Vanderbilt, under the supervision of Rendigs Fels,¹ combines the case method of instruction² with the self-paced personalized system of instruction (PSI) developed by Fred S. Keller [5].

The course consists of twenty lessons. Each lesson includes a short examination on basic concepts (seven exams), analytical skills (seven exams), or policy cases (six exams). The course has no lectures. Students are given assignments, study them with the help of a proctor, and take the examinations when they think they are prepared. The criterion for passing is mastery—one hundred percent. Students who do not meet the criterion are recycled, and, after additional study may retake tests on lessons they fail to master. This process continues until each lesson is passed or the semester ends. Grades depend partly on how many lessons are completed during the semester and partly on a final examination. The case method is in-

tegrated into the PSI method in the final six lessons. In these lessons students analyze realistic policy issues in a systematic way.

There has been a substantial amount of research on PSI courses summarized in [6]. Most of this research has been conducted in psychology classes. Elizabeth Allison has recently reviewed the application of PSI principles to college level economic instruction [1]. Scott McCuskey has evaluated an adaptation of the Vanderbilt-JCEE course in his Ph.D. thesis [7].

The results of evaluations of student learning in PSI courses are inconclusive, although PSI students seem to do at least as well as conventionally instructed students on standardized multiple-choice examinations. In general students report that they enjoy PSI courses more than traditional courses and usually think that they have learned more than they would have in a conventional course (but the objective data do not always support this hypothesis). Instructors commonly report that it is exciting to teach a PSI course and there is some scattered evidence that PSI courses produce a higher percentage of students concentrating in the field of the course.

The economics of PSI courses has not been examined as carefully as their educational value. PSI courses require significant start-up costs plus substantial instructional resources during their operation. The common organization of PSI courses has one instructor supervising a group of proctors, who in turn each tutor up to ten students. Thus for a course of one hundred students there would be at least eleven teachers. This is generally beyond the cost capabilities of most colleges and universities. For this reason, and because many faculty believe that

* This evaluation is part of a larger project evaluating the personalized self-paced case-method of instruction of elementary economics being conducted experimentally by Vanderbilt University and the Joint Council for Economic Education. Financial support was received from the JCEE, the Ford Foundation, and Vanderbilt. Lee Webby, Mike Cook, Tom Overstreet, and Martin Rini provided clerical assistance. Rendigs Fels and Stephen Strand contributed to the project from its inception to final reporting of the results.

¹ An extensive description of the course, including all materials necessary to replicate it, is contained in [3].

² The essence of the case method is requiring students to systematically think through real-world problems for themselves. A discussion of the case method is included in [2]. Many of the cases used in the course are reported in [4].

one learns best when one is instructing a course, most PSI courses have utilized undergraduate students as proctors. Only in this way can the course be made cost-effective.

Allison says that after the course has been developed it requires less time from the instructor than a conventional lecture course [1, 8]. Fels contends that the supervision of the PSI course plus the supervision of the proctors means that such courses place a heavier burden on the instructor than comparable conventionally taught courses [3, 10]. He argues that only if the instructor is given teaching credit for the proctors as well as for the students enrolled in the PSI course do the benefits exceed the costs to the instructor. Some schools reward the undergraduate students with course credit; others compensate the proctors financially. Whether the instructor is given teaching credit for the proctors is intimately connected with the form of compensation received by the proctors.

If teaching credit for supervising the proctors and compensating proctors with course credit rather than cash are necessary to make the course cost-effective to both the school and the instructor, it is critical that the educational content of the proctoring experience be evaluated. While PSI courses commonly use students as proctors, and many assertions are made that the student-proctor learns more from this activity than available alternatives, there have been no systematic empirical efforts to validate these claims.

This study is an attempt to evaluate the costs and benefits that accrue to proctors in the Vanderbilt-JCEE PSI case-method course.² Twenty-one proctors were employed during the 1974-1975 academic year. They answered students' questions, administered tests, and gave provisional grades to tests and papers. Each proctor was responsible for about four to five students. The proctors were all undergraduate students, primarily

juniors and seniors with good grade records. Most of them had previously taken intermediate microeconomics and intermediate macroeconomics. The proctors took each of the examinations themselves from Fels prior to administering them to enrolled students. They were compensated for their services with three semester hours of credit.

Each proctor was interviewed immediately after completing the proctoring experience. In addition, examination scores for the spring proctors and a control group are analyzed. Multiple linear regression analysis is used to examine the hypothesis that students develop a more complete understanding of material by instructing others. The evidence supports that hypothesis.

II. INTERVIEW RESULTS

All twenty-one proctors thought that they learned more economics by proctoring than they would have learned in an upper level economics course. Seventeen of the twenty-one thought that academic credit was the appropriate form of compensation for proctoring. The strongest argument against awarding academic credit for proctoring is that the economic principles learned while proctoring should have been learned when the proctors took the elementary course themselves. The counter-argument is that it is not necessary to master anything close to one hundred percent of the material presented to students in conventional elementary economics classes in order to earn an A-. An A- level of understanding is far below that necessary to teach the principles to other students. The academic credit for proctoring is for learning "new material" with which the proctors were previously familiar, but which they did not completely understand. The academic credit is for changing "familiarity" into "understanding" of much of the material in the elementary course.

The student-proctors suggested many benefits from the experience that extend beyond their improved understanding of economic

² A complete evaluation of the Vanderbilt-JCEE course on additional criteria is contained in [8].

principles. They commonly cited an improvement in the clarity of their verbal expression, insight into the teaching process and how people learn things, experience they obtained in motivating people, and the interesting (and sometimes trying) experience of recycling people.

Fourteen of the twenty-one proctors identified time as the predominant cost to them of proctoring. Many of the proctors claimed that proctoring took about twice as much time as studying for an alternative upper level economics course. Several proctors complained about the distribution of time required. They apparently were frustrated by their lack of control over their personal time. They had to take the examinations themselves at least as rapidly as their most enthusiastic student and be available to students when they needed help, which sometimes interfered with the proctors' study plans for their other courses.

There were very few costs identified by proctors other than time. A few mentioned inevitable personality conflicts. The responsibility for other people's progress eliminates proctors' options to "just go for a C." Some proctors indicated that recycling students caused them some emotional strain. Only one proctor thought that proctoring hurt his grades in other courses that he was taking simultaneously. Most of the proctors recognized a trade-off between time and pressure. They perceived less pressure on them as proctors than there would have been in an upper level economics course. There were fewer crises (i.e., exams). One proctor put it bluntly: "more time, but less sweat."

III. CONTROLLED EXPERIMENT

An empirical evaluation of the impact of proctoring on understanding economic principles was conducted in spring 1975. To control for initial understanding a test of economic principles was administered to ten proctors and twenty matched students prior to the semester. All of these students took the same examination at the end of the se-

mester, after all classes had been completed, but prior to final examinations.

The test instrument was the new (1975) introductory economics test (micro and macro) in the College Level Examination Program (CLEP) of the Educational Testing Service. This is a 100 item 90 minute multiple-choice test of economic principles. It stresses understanding of abstract economic theory. It is a difficult examination of superior quality and thus was particularly suitable for our purpose, since we wished to discriminate among different levels of understanding at relatively high levels of competence. However, the CLEP examination does not test skill in applications, which constitutes about one-fourth of the experimental course. All students were paid a flat fee of ten dollars to take the two exams. Students were motivated to perform well on the CLEP examination by offering cash prizes of up to ten dollars to the top fifteen scorers on each exam, the size of the prize being related to performance.

The ten proctors were selected by Fels prior to the spring semester. The twenty control group students were matched two to a proctor on the criteria of cumulative grade point average, year in school, background in intermediate microeconomics and intermediate macroeconomics, major, and number of previous economics courses taken. To test the similarity of the control and experimental groups we compared their performance on the CLEP examination before the semester. The proctors averaged 72.4, while the control group averaged 72.6. There is no statistically significant difference between these means. On the basis of this evidence, and considering our matching procedure, we concluded that the experimental and control groups originated from the same population and thus could be used in pooled regression analysis.

The index of performance measuring improved understanding of economics principles during the semester is the difference between the CLEP score after the semester and the CLEP score before the semester.

This is a value added measure; it controls for students' initial understanding of economics principles. Since the goal is to relate improvement in understanding of economics principles *during the semester of proctoring* to the proctoring activity, a value added measure is most appropriate. However, because students who perform particularly well on the pre-test cannot score very well on a value added measure we used an alternative measure of improvement in understanding of economics principles, value added divided by the number of points that were available to be added during the semester.⁴ Symbolically the measures of performance are:

$$P_1 = T_2 - T_1, \quad \text{and} \quad P_2 = \frac{T_2 - T_1}{100 - T_1},$$

where T_i is the score on the i th test and P_j is the j th performance index.

A simple comparison of mean performance between the two groups reveals that the proctors did appreciably better than the control group. These results are reported in Table I. The proctors' mean is significantly higher than the conventional students' mean at the 0.05 significance level. The pre-test scores are essentially identical. On the average the proctors gained 9.4 points during the semester while the control group gained 3.5 points.

There may have been significant differences in the learning experience, environment, or background of students that are not controlled in a simple comparison of mean scores. In particular, different students take different numbers of courses. In addition, it may be important to see if there are systematic differences in the other standard factors that might explain the difference in performance between proctors and control students. We expect most of the control variables like grade point average, previous intermediate theory courses, cumulative number of economics courses, and year in school to show little impact on differential performance because we initially matched the proctor and

control samples on the basis of these criteria. Other control variables include SAT scores, sex, mathematics background, other workload during the semester, and number of other economics courses during the semester. These variables have all been examined in much detail in the literature of economics education and most of the hypotheses relating them to learning are self-evident. Therefore the results of the multiple linear regressions are reported in Table II without further elaboration.

The regression results indicate that only proctoring and the number of economics courses that a student took during the experimental semester had a significant effect on P_1 . Not one of the remaining control variables was statistically significant at the .05 level. This provides support for the contention that proctoring in the PSI elementary course provides substantial improvement in the understanding of economic principles by the proctors. In addition, it indicates that alternative economic courses are of value for achieving the same goal. It is somewhat comforting to discover that the major device universities employ to improve student understanding demonstrates a statistically significant impact in the expected direction and that other environmental and socio-demographic variables do not, by themselves, explain much of the difference in improved understanding among students.

Because we expected proctoring and alternative economics courses to improve performance scores, these variables are subjected to one-tail statistical significance tests. Age, sex, SAT scores, semester hours, and total economics courses are subjected to a two-tail statistical significance test because there are competing hypotheses with respect to the predicted sign of the coefficient of each of these variables. In additional regression models a dummy variable for whether a student had taken calculus, students' grade point average, and a series of dummy variables for whether students had taken intermediate micro-economic theory, intermediate macroeconomic theory, and the

⁴ This measure was, to my knowledge, first proposed by Whitney [9].

COMMUNICATIONS

TABLE I
MEAN PERFORMANCE OF PROCTORS AND CONTROL STUDENTS ON CLEP EXAMINATION

Performance	Proctors	Control Students	t-ratio for difference between means
T ₁	72.4	72.6	0.4
P ₁	+9.4	+3.5	2.82**
P ₂	+0.333	+0.155	2.20**
n	10	20	

** = statistically significant at .01 level.

grades they obtained in the intermediate theory courses proved to be statistically non-significant. This result was expected since students were matched in the two samples on the basis of grade point average and intermediate economic theory experience. Since the presence of the other control variables did not affect the behavior of the coefficients or standard errors of the variables included in regression (2), the details of these other regressions are not reported.

Because the number of current semester courses in economics is not significant in regression (1), the coefficient should be interpreted as zero. Thus a comparison between proctoring and taking advanced economics courses will obviously favor proctoring in regression (1), since there is no impact of advanced economics courses. However, it is appropriate to control for those factors that were not considered in the matching procedure. Therefore, regression (2) should be examined.

The results of regression (2) indicate that an additional economics course during the semester is associated with a 7.8 percent increase in percent of potential value added to the CLEP score. Proctoring is associated with a 20.8 percent increase in percent of potential value added achieved. According to these figures, proctoring is associated with

2.7 times the increase in CLEP performance during the semester than would result from taking one advanced economics course. If the relative cost to the students of these two alternatives is two to one (many proctors stated that the time demand of proctoring was twice that of an advanced course), then in terms of the test criterion proctoring seems more favorable than an advanced economics course by a ratio of about three to two.

The analysis indicates that course instruction and proctoring were the only significant determinants of how much economic theory was learned by upper-class students. However, several caveats are in order. First, these conclusions come from a relatively small sample of students in one discipline at one university. Whether they can be applied to other educational environments is problematical. On the other hand, these empirical data are, to our knowledge, the first systematic attempt to test the thesis that the most effective way to learn economics is to instruct it. In an area with an abundance of ad hoc theorizing and self-appointed experts, bringing systematic empirical data to bear on the question must be considered progress. Second, only linear models were tested. For example, we did not consider whether the impact of an advanced economics course on

COMMUNICATIONS

TABLE II
MULTIPLE REGRESSION RESULTS: DEPENDENT VARIABLE IS (POSTTEST-PRETEST)/(100-PRETEST)

Variable	(1)	(2)	Measure, [Mean]
Proctor	.158* (1.95)	.208* (2.40)	proctor = 1, otherwise = 0, [.333]
Concurrent Economics Courses	.038 (0.99)	.078* (1.81)	number of spring 1975 economics courses, [2.87]
Age		-.012 (-1.05)	years (proxy for year in school), [19.67]
Sex		.073 (0.70)	male = 1, female = 0, [.70]
SAT verbal		.0016 (1.69)	numerical score, [669]
SAT quantitative		.0004 (0.51)	numerical score, [579]
Semester hours		-.021 (-1.24)	semester credit hours in spring 1975, [14.1]
Cumulative Economics Courses		.006 (0.48)	cumulative economics courses as of June 1975, [10.5]
R ² (coefficient of determination)	.183	.497	
F (Fisher's F-ratio)	3.0*	2.5*	

n = 30

* = statistically significant at 0.05 level.

improving the understanding of economic theory depends on whether it is the first, second, third, or fourth economics course taken during the semester. Some people would argue that there is a threshold effect, that is, a minimum number of courses necessary to have any impact at all. Others would argue that there are diminishing marginal returns to improving one's understanding of economic theory with respect to additional courses. A non-linear impact might also be expected from several of the other variables in our regressions. Alternative functional forms have not been examined empirically because of the limited size of the sample.

Third, the empirical test explored the relative effect of proctoring and advanced courses only on improving students' understanding of economic theory. There are other worthwhile goals of advanced economics courses—providing familiarity with institutions, generating enthusiasm for economic inquiry, exposing students to research methods, etc. None of these objectives are measured effectively by the test instrument. On the other hand, there are also many benefits of proctoring that go beyond the improved understanding of economic theory. Proctors presumably improve their skills in analyzing realistic policy cases; they learn to

deal with other people in conflict situations; they are forced into self discipline and develop patience. These factors may easily be as important to students' future life experience as learning about economic institutions or research methodology. We have evaluated only one dimension of the multi-dimensional output of university instruction. This comparison can be useful if it permits an objective evaluation of alternative instructional techniques on this single dimension. Then more effort can be directed toward comparative evaluation of the other dimensions of instruction and the all important problems of how these different dimensions should be weighted in instructional decisions.

IV. CONCLUSION

Based on interviews of proctors and on empirical analysis, it is concluded that proctoring is an effective means for students to improve their understanding of economic principles. The twenty-one proctors during 1974-1975 unanimously voiced the view that they learned more economics by proctoring than by taking an alternative advanced economics course. An empirical study of ten proctors and twenty control students indicates that the proctors would improve their scores on an economics exam 2.7 times the improvement by students who elected to take an advanced economics course instead. On the basis of this evidence an argument can be made for compensating proctors for their services with academic credit.

Studies which conclude that students in PSI courses perform no differently from students in conventional lecture-discussion courses have failed to enumerate an impor-

tant set of benefits that accrue to the student-proctors. This supports a case for employing students (rather than professional instructors) as proctors for PSI courses. If student-proctors are compensated with course credit and the instructor is given teaching credit for the proctors, PSI courses stand a good chance of being cost effective. The evidence uncovered in this analysis suggests that there is more learning of economic theory taking place during proctoring than during alternative economics courses, which provides a logical foundation for compensating undergraduate student-proctors in PSI courses with academic credit.

JOHN J. SIEGFRIED

Vanderbilt University

REFERENCES

1. Allison, E., "Self-Paced Instruction: A Review," *The Journal of Economic Education*, Fall 1975, 5-12.
2. Fels, R., "Developing Independent Problem-Solving Ability in Elementary Economics," *American Economic Review*, May 1974, 403-407.
3. ———, "The Vanderbilt-JCEE Experimental Course in Elementary Economics," *The Journal of Economic Education*, Winter 1974, Special Issue No. 2.
4. ——— and Uhler, R.G. *Casebook of Economic Problems and Policies: Practice in Thinking*. St. Paul, Minn.: West Publishing Company, 1975.
5. Keller, F.S., "Good-bye, Teacher!" *Journal of Applied Behavioral Analysis*, 1968, 79.
6. Kulik, J.A., Kulik, C.L. and Carmichael, K., "The Keller Plan in Science Teaching," *Science*, February 1, 1974, 379-383.
7. McCuskey, E.S., "Evaluation of Modified Versions of the Vanderbilt-JCEE Experimental Course in Practice." Ph.D. Dissertation, Nova University, 1975.
8. Siegfried, J. and Strand S., "An Evaluation of the Vanderbilt-JCEE Experimental Course in Elementary Economics." Unpublished manuscript, Vanderbilt University, 1976.
9. Whitney, S., "Measuring the Success of the Elementary Course," *American Economic Review*, March 1960, 159-169.

Teacher Effectiveness and Student Performance

Howard P. Tuckman

This paper evaluates the effects of using graduate instructors rather than experienced faculty in the macroeconomics principles course. Because a single measure of teacher effectiveness may cause an increase in "measured" rather than "desired" output [1], we employ several measures of effectiveness. The first (Q_1) is student score on a ten-question version of the Test of Understanding in College Economics (TUCE).¹ The second (Q_2) is student score on a twenty-question economic attitudes (AS) test, constructed along the lines suggested by Mann and Fusfeld (MF) [4]. Measures of this type are not free of value judgments and do not allow for the theoretical divisions which currently exist within the profession. However, economic attitudes are changed in the principles class and an effort should be made to measure the change, as MF found that an economic attitudes test captures learning changes in subject matter-oriented students while grades are better measures of effectiveness in teaching success-oriented students.

Final grade (Q_3) for the quarter ($A = 4, B = 3, \dots, D = 1$) is the third measure of effectiveness. A student's grade reflects the instructor's judgment of how well he performs. If some instructors differ substantially in their grading procedures or in their ability to teach, this has a bearing on their effectiveness and is germane to our study. The fourth measure (Q_4) is an index of student interest in economics ($5 = \text{very high}, 4 = \text{high}, \dots, 1 = \text{very low}$) based on a set of questions asked to students at the end of the course. The final measure (Q_5) is a measure of student willingness to take another course in economics ($1 = \text{student intends to take another course}, 0 = \text{otherwise}$). The latter two measures are intended to capture the instructor's ability to interest students in the materials.

Zero-order correlation coefficients for each measure are shown in Table 1. Note that the largest correlation (between AS and TUCE) is only 0.28. Moreover, a consistent pattern emerges; the correlations between the AS and other measures are generally higher than between either the TUCE and other measures or the final grade and other measures.

Data Description

Beginning in the 1972-73 winter quarter and for five successive quarters thereafter, students in the macroeconomics principles courses at Florida State were asked to complete two exams (TUCE and AS) and a detailed questionnaire at both the beginning and end of the quarter. A total of 612 students were tested.² Of these 548 or about 90 percent provided usable data.³ A total of 12 classes were included in the sample, taught by one full professor, two associates, two assistants, and three graduate instructors. On the basis of prior studies in this area the following independent variables were utilized: X_1 = grade point average ($5 = 3.5-4.0$,

Howard P. Tuckman is Associate Professor at the Institute for Social Research, The Florida State University.

Table 1
Zero-Order Correlations Among the Dependent Variables

	TUCE	AS	Final Grade	Interest in Economics	Continuation
TUCE	—	0.28	-.07	-.12	.01
AS		—	.20	-.17	.07
Final grade			—	.07	.04
Interest				—	.13
Continuation					—

4 = 3.49-3.0, . . . , 1 = 1.99 or less), X_2 = score on TUCE pretest, X_3 = score on AS pretest, X_4 = sex (1 = female, 0 = otherwise), X_5 = major (1 = economics or natural science, 0 = otherwise), X_6 = precourse interest in economics (5 = very high, 4 = high, . . . , 1 = very low), X_7 = class standing (5 = graduate student, 4 = senior, . . . , 1 = freshman), X_8 = precourse willingness to take another nonrequired course in economics (1 = if willing, 0 = otherwise), Y_1 = number of years since instructor received his Ph.D. (A third-year graduate student is two years away from receiving his Ph.D. and thus enters the regression with a -2), Y_2 = number of hours taught by the instructor during the quarter, Y_3 = graduate instructor (1 = if a graduate student, 0 = otherwise).

The Role of Experienced Teachers

Several studies have shown that teacher experience affects student performance. High school students taught by experienced teachers receive higher scores on national tests than those taught by less experienced teachers [3]. Likewise, experienced teachers produce fewer dropouts and more students who wish to pursue their education [6]. While the evidence for college teachers is less extensive, it appears that a teacher's experience affects the performance of his or her students.

Regression analysis is used to examine the relationship between teacher experience and student performance, controlling for the effects of student and other nonteacher-related characteristics. Each column of Table 2 gives a regression equation for one of the Q_1 outputs; each row shows the regression coefficients for one of the independent variables, along with its t -value (in parentheses below the coefficient⁴). How does teacher experience affect student performance? The years beyond Ph.D. coefficient (Y_1) is statistically significant only in the TUCE and AS equations. At the sample mean of 7.6 years of experience, the average post-TUCE score is raised by 0.3 point or 7 percent above the pre-TUCE mean score. Post-AS score rises by 0.65 point or about 6 percent above the pre-AS mean score. These computations do not suggest that experienced teachers contribute a great deal to student learning.⁵ In the post-TUCE equation, the effect of one year of additional teaching experience (0.039) is substantially less than that of increasing the average grade point of students by 0.1 point (0.054).⁶ This seems to imply that student rather than instructor quality is the key to classroom performance.⁷

Consider the other coefficients in Table 2:

1. Note that the constant term in each of the regressions is positive and significant. This suggests that other variables which have been excluded from the equation may be significant. Thus far, the absence of an effective learning theory has prevented researchers from identifying other determinants of learning.
2. The positive coefficient on the grade point average variable suggests that students who do well in other courses also do well in economics. When this finding is combined with the lack of significance for the Pre-Interest, Pre-Continue, and student-major variables it may

Table 2
The Effect of Teacher Years of Experience
as Indicated by Several Performance Measures

Independent Variables	Post-TUCE	Post-AS	Final Grade	Interest in Economics	Continue in Economics
Constant term	5.591 (16.6)	12.760 (18.4)	3.719 (14.5)	1.543 (9.0)	0.360 (6.0)
Grade point (X_1)	0.542 (6.2)	0.631 (5.6)	0.376 (9.7)	—	—
Pre-TUCE (X_2)	0.219 (5.5)	—	—	—	—
Pre-AS (X_3)	—	0.313 (9.5)	0.050 (4.4)	—	—
Sex (X_4)	-0.371 (2.1)	—	—	-0.306 (2.8)	-0.109 (3.1)
Major (X_5)	—	—	—	—	*
Pre-Interest (X_6)	—	—	—	0.347 (5.7)	0.058 (2.9)
Class Standing (X_7)	—	—	-0.1190 (2.5)	—	—
Pre-Continue (X_8)	—	—	—	—	0.408 (8.5)
Years beyond Ph.D. (Y_1)	0.039 (4.0)	0.085 (3.7)	—	—	—
Number of hours taught (Y_2)	—	-0.462 (5.1)	—	—	—
Adjusted R^2	0.18	0.25	0.22	0.08	0.21
f -Ratio	27.11	40.94	33.27	22.68	31.95
Adjusted standard error	1.87	2.44	0.84	1.17	0.38

*Significant at 5-percent level if Pre-Continue variable is eliminated but insignificant when it is included.

- suggest that a student's learning skill, rather than his prior interest in economics, is the crucial factor in determining how well he learns economic principles.
3. A student's pre-TUCE score presumably reflects his knowledge of economics when he enters the course. Surprisingly, a high prescore does not appear related to a high final grade, again suggesting that prior exposure to the field may not be as important as the ability to learn.
 4. Performance on the pre-AS exam is significant in determining post-AS score and final grade. Since the AS test presumably measures economic *reasoning* rather than prior knowledge, this may explain why AS is significant and TUCE is not. An increase of about 3.3 points on the pre-AS exam adds one point to the post-AS score and 0.17 points to the final grade.
 5. Women perform less well on the TUCE than do males. They are also less interested in economics both when they enter and when they leave the course. This is reflected in fewer female continuations to other economic courses.
 6. We know surprisingly little about what determines a student's interest in economics, either before or after the course.
 7. The hours taught variable is included in the regression to control for the effects of variation in workload on teacher effectiveness. It is difficult to interpret in a small sample, however,

since hours taught are not evenly distributed across ranks. An increase in the number of hours taught appears to have an effect on post-AS score but this may be due to its correlation with the graduate instructor variable.

8. Finally, although the $\bar{R}^2$'s are low, the f tests are all significant at a 1-percent level and the $\bar{R}^2$'s are in line with those of other researchers in this area.

The Effectiveness of Graduate Instructors

Since experienced teachers raise student scores on the TUCE and AS, one might expect that departments using graduate instructors in lieu of faculty experience a decline in student learning. Such an expectation ignores the possibility that graduate instructors compensate for their lack of experience by their enthusiasm, efforts to identify what their students don't understand, approachability, and greater rapport with their class. To determine the effect of graduate instructors in the principles sequence during the summer of 1973 and in subsequent quarters, the economics department offered several sections taught by graduate students. These students developed their lectures free from departmental directive, although they followed a common course outline.⁹ The sole criterion for selection as an instructor involved the student's past rating as a teaching assistant.

If graduate instructors affect student performance, this should be reflected in a significant t -value for the Y_3 variable when it is included in the regressions discussed above. Estimate I in Table 3 shows the regression coefficient obtained by including Y_3 in the above regressions. Only the regression coefficients for Y_1 and Y_3 and the adjusted R^2 and f -statistic are shown since most of the other regression coefficients do not change substantially. Because instructors teach fewer hours, however, Y_3 is correlated with Y_2 . Likewise, Y_3 and Y_1 are also correlated. Thus, the inclusion of all three variables masks the significance of the Y_3 variable. To provide an alternative test of the effect of graduate instructors, the Y_1 and Y_2 variables were removed from the model and the regressions rerun with only the Y_3 variable (Estimate II).

Table 3
Two Estimates of the Effect of Graduate Instructors
on Student Performance

	Post-TUCE	Post-AS	Final Grade	Interest in Economics	Continue in Economics
<i>Graduate Instructor</i>					
Estimate I	0.193 (0.7)	0.024 (0.1)	-0.311 (1.9)	0.214 (1.6)	0.014 (0.3)
Estimate II	-0.362 (1.7)	1.011 (3.6)	-0.014 (0.2)	0.214 (1.6)	0.014 (0.3)
<i>Years beyond Ph.D.</i>					
Estimate I	0.044 (3.7)	0.085 (3.5)	*	*	*
<i>Adjusted R^2</i>					
Estimate I	0.18	0.25	0.22	0.09	0.21
Estimate II	0.16	0.23	0.21	0.09	0.21
<i>f-Ratio</i>					
Estimate I	21.78	32.68	27.53	16.07	25.53
Estimate II	23.09	48.57	32.76	16.06	25.53

*Denotes that the variable was not included in Table 3 because it lacked statistical significance.

The specification of the equation affects our conclusions regarding the effects of graduate instructors. With Y_3 included as the sole instructor variable, the results suggest that graduate instructors have a positive effect on student performance on the AS test and a negative effect on

performance on the TUCE. They have a nonsignificant effect on the other measures. With the three Y variables included together, Y_3 has no significant effect on TUCE and AS but a negative effect on the final grade.⁹ Taken together the results suggest that experienced faculty may do a better job in teaching the skills required for the TUCE exams while graduate instructors have a more positive effect on student attitudes.

Conclusion

Prior studies raise a paradox highlighted in our findings. Experienced faculty presumably have a positive effect on student performance yet graduate instructors appear to be as effective as faculty in teaching economic principles [5]. Can this paradox be explained? What may be involved here are different sets of skills: in the case of graduate students, availability to grasp what students don't understand, enthusiasm, and approachability; for experienced faculty, a greater depth of understanding of the technical material, greater self-confidence, a more critical approach to the subject. Given the approach followed in this study, one can only speculate as to the underlying factors explaining our results. The measures of learning used here capture only part of the total learning environment. An effective instructor conveys many things not captured by our measures: excitement with the subject, caution in accepting unsupported arguments, a perspective on the economic system, etc. We have yet to develop measures which provide an effective comparison of faculty and graduate instructors along these lines. Given the mixed nature of our findings, however, we would urge economic researchers to use more than one measure of effectiveness in studies of this type.

Footnotes

¹We were forced to limit the number of questions asked because of time constraints. Previous studies in this journal found that TUCE questions have a high degree of internal reliability. The Kuder-Richardson formula 20 statistic for the modified exam was 0.6.

²The total enrollment in the period was over 1,000. Since it was not possible to test all sections taught in the department, a representative sampling of the principal faculty was obtained.

³Two factors account for the less than 100 percent response rate: an inability to match pre- and postquestionnaire and absence when either the pre- or posttest was given. A check of the unmatched data suggests that no bias is introduced from the former source. Every effort was made to eliminate absences as the study progressed. Whether they introduce a serious source of bias is uncertain.

⁴Only the significant variables are included in the final regression equation.

⁵On the point see [8].

⁶Obviously this option is not open to most departments.

⁷Ideally, years of teaching experience rather than years beyond Ph.D. should have been used. For the faculty in this sample, the distinction is not important. Likewise, because of the distribution of course loads it was not necessary to control for whether an instructor previously taught introductory economics or whether his major assignment was advanced undergraduate or graduate level courses.

⁸The exception to this rule was the fall quarter where one faculty member and two graduate instructors taught a coordinated set of courses. See Barbara and Howard Tuckman, "Toward a More Effective Economic Principles Class," *Journal of Economic Education*, Special Issue No. 3 (Spring 1975).

⁹The insignificant coefficient for Y_3 in Q_1 of estimate I seems to be because Y_1 and Y_2 pick up this variation when they are included in the regression. The negative coefficient for Y_3 in the Q_3 equation appears to be due to a difference in the grading procedures followed by two faculty in the sample. See Barbara and Howard Tuckman, *op. cit.*

REFERENCES

1. A. Kelley, "Uses and Abuses of Course Evaluations as Measures of Educational Output," *Journal of Economic Education*, 4 (Fall 1972).
2. C. Lamphear and C. McConnell, "A Note on the Use of Graduate Teaching Assistants in the Principles Course," *Journal of Economic Education*, 1 (Spring 1970).

3. H. Levin, "New Models of School Effectiveness," *Do Teachers Make a Difference?* Washington: G.P.O., 1971.
4. W. Mann and D. Fusfeld, "Attitude Sophistication and Effective Teaching in Economics," *Journal of Economic Education*, 1 (Spring 1970).
5. W. Oates and R. Quandt, "The Effectiveness of Graduate Students as Teachers of the Principles of Economics," *Journal of Economic Education*, 1 (Spring 1970).
6. H. Tuckman, "High School Inputs and Their Contribution to School Performance," *The Journal of Human Resources* (Fall 1971).
7. D. Weidenaar and J. Dodson, Jr., "The Effectiveness of Economics Instruction in Two-Year Colleges," *Journal of Economic Education*, 4 (Fall 1972).
8. A. Welsh and R. Fels, "Performance on the New Test of Understanding in College Economics," *American Economic Review*, 59 (May 1969).

Kim Sosin and Campbell R. McConnell

The Impact of Introductory Economics on Student Perceptions of Income Distribution

The purposes of this paper are (1) to survey student attitudes on the issue of income inequality; (2) to determine whether one semester of economics causes any statistically significant shifts in student attitudes on the income distribution issue; and (3) to determine which, if any, of a number of student background characteristics might be significant in explaining student perceptions of what the personal distribution of income ought to be.

Design of the Study

The primary focus of this study was an experimental group consisting of students at the University of Nebraska-Lincoln, mostly sophomores, who were enrolled in the first semester of a two-semester principles-of-economics sequence. The control group was made up of students enrolled in an introductory human geography course. The only function of the control group was to permit a test of the proposition that changes in attitudes of economics students occur because of the study of economics rather than external events. Students in the control group were dropped from the study if they had taken the economics course, since it might have influenced their preassessment attitudes.

During the semester students in the experimental group covered a range of subject matter generally associated with the macro segment of the principles course. The semester's work centered upon four areas: (1) an introductory segment which defines economics and surveys basic concepts and institutions (e.g., markets and prices, comparative advantage, the role of the public sector, etc.); (2) macro theory and fiscal policy; (3) money, banking, and monetary policy; and (4) economic growth. In connection with a brief analysis of the redistributive function of government, students were assigned reading material which surveys basic empirical data on personal income distribution, discusses the causes of income inequality, and presents the various arguments which comprise the cases for and against greater income equality (specifically, see McConnell, pp. 108, 760-767). Although the income distribution issue was not explored at great length in class lectures, the actual distribution of personal income by quintiles was presented and it was noted that (1) certain groups would not be able to participate effectively in a market economy and hence would receive little or no earned income and (2) government has in fact attempted by a variety of means (direct market intervention, welfare programs, and tax policies) to

Kim Sosin is assistant professor of economics and Campbell R. McConnell is professor of economics at the University of Nebraska-Lincoln.

TABLE 1
Attitude Assessment Scores^a

Group	Mean	Standard Deviation
Experimental (economics) group (<i>n</i> = 134) ^b		
Preassessment	-3.4104	4.7969
Postassessment	-0.7388	5.0604
Individual difference	2.6716	4.1273
Control (geography) group (<i>n</i> = 38) ^c		
Preassessment	-1.5526	5.1133
Postassessment	-0.5526	4.6771
Individual difference	1.000	3.8765

^aA positive (negative) score indicates a preference for income equality (inequality).

^bAlthough 191 students were enrolled, only 134 were available for both pre- and post-testing.

^cOf the 52 students in geography, 38 took both the pre- and post-test.

redistribute income. Following Arthur Okun, the debate over income distribution was summarized as an equality-efficiency trade-off problem, complete with his "leaky-bucket" analogy.¹ In addition, many other segments of the course touched upon or implied income (re)distribution issues, e.g., unemployment and inflation, distribution of the tax burden, international income differentials, etc. In presenting the income distribution issue the instructor was cautious *not* to reveal his personal prejudices nor to communicate material in such a fashion as to alter student values in a particular direction.

The measurement instrument used in this study was an attitudinal survey which the authors prepared and administered at the beginning and end of the semester. This survey consisted of eleven statements pertaining to income distribution rules and policies. Each student was asked to indicate whether he or she "strongly agreed" (SA), "agreed" (A), was "undecided or neutral" (U), "disagreed" (D), or "strongly disagreed" (SD) with each statement. For all statements which favored (greater) income equality numerical weights of +2, +1, 0, -1, and -2 were assigned to SA, A, U, D, and SD responses respectively. Those weights were reversed for those statements favoring (greater) income *inequality*. In other words, although the responses are ordinal, they were treated as cardinal to yield an aggregate numerical score. Possible scores could range from -22, indicating an extremely strong preference for income inequality, to +22, reflecting the strongest possible egalitarian position.

At the start of the semester a variety of background information was requested of each student, including sex, race, rural-urban character of permanent residence, marital status, age, cumulative grade average, hours of employment, and occupational-income status of parents. To foster anonymity and thereby response accuracy, all surveys were identified by partial social security numbers rather than names.

Attitudes and Attitudinal Shifts

Mean scores of the experimental and control groups are presented in Table 1 on a preassessment (beginning of semester) and postassessment (end of semester) basis. We observe that the stance of both groups was somewhat nonegalitarian at the start of the semester, the economics group being more so than the geography group. Both groups shifted to a less nonegalitarian stance by the end of the semester. A reasonable interpretation of the rather pronounced shift of the economics group might be that the material encountered in the course made students more open-minded and therefore less certain of their original stance.

Are the changes in scores significant within each of the two groups? Calculation of t values indicated that the change for the economics group was highly significant at the .01 level ($t = -7.4930$). The change for the geography control group, however, was not significant ($t = -1.5902$). Is the attitudinal change for economic students significantly different from the change for the geography students? The calculated t value ($t = -2.232$) proved to be significant at the .05 level. The implication of these results is that exposure to a rather typical semester of macroeconomics does seem to entail a statistically significant attitudinal shift toward egalitarianism.

It is of some interest to compare the responses in each group to determine the directions of the individual attitudinal changes. Are the differences in preassessment and postassessment responses among students consistent with the hypothesis that only random events influenced their attitudinal development over the semester or is there evidence of systematic influences? If events were random in impact, our results would be consistent with a hypothesis that 50 percent of the students became more egalitarian and 50 percent became less, i.e., a population proportion of .50. In the economics group, 67.2 percent became more egalitarian, which is significantly different from 50 percent ($t = 4.241$). In contrast, in the geography group, 52.6 percent became more egalitarian, but that figure is not significantly different from 50 percent ($t = .642$). The conclusion is that the economics course had a systematic influence on the direction of student attitudes.

Explaining Attitudes: Model and Findings

A multiple regression model was formulated to explain the attitudes of beginning economics students toward income distribution in terms of student background characteristics. The model, which employs the preassessment scores of the economics group ($n = 191$),² is of the following form:

$$ATT = a_0 + a_1S + a_2R + a_3P + a_4M + a_5G + a_6H + a_7A + a_8O + a_9I + u$$

where:

ATT = student attitudes on income distribution (attitude survey score)

S = sex (dummy variable: male = 0; female = 1)

R = race (dummy variable: white = 0; all others = 1)

P = population of permanent residence (dummy variable: rural, defined as living on a farm or in a town under 2,000 = 0; urban, defined as living in a town or city of 2,000 or more = 1)

M = marital status (dummy variable: single = 0; married = 1)

TABLE 2
Hypotheses, Rationales, and Expected Signs
for Independent Variables

Variable and Hypothesis	Rationale	Expected Sign
<i>S</i> Females are more egalitarian than males.	Anticipation of a work-life cycle involving periods of dependence upon intrahousehold transfers will cause females to be more favorably disposed toward income redistribution.	+
<i>R</i> "Others" will be more egalitarian than whites.	As compared to whites, there is a greater probability that "others" (nonwhites) come from low-income backgrounds and have been beneficiaries of income transfers.	+
<i>P</i> Students with rural backgrounds are more egalitarian than those with urban backgrounds.	Historically, the rural (agricultural) sector has benefited from income transfers. Therefore, students from rural areas will be more favorably disposed toward redistributive programs.	-
<i>M</i> Married students will be more egalitarian than single students.	Intrahousehold transfers of income are a basic characteristic of households, e.g., the wife may be working to put her husband through college; hence there is an immediate awareness of the potential benefits of income transfers.	+
<i>G</i> The higher a student's grade average, the less egalitarian he/she will be.	Students who perform well academically will have reason to believe they will also perform well in the economy and therefore will not be beneficiaries of egalitarian policies.	-
<i>H</i> The more hours a student works per week, the less egalitarian he/she will be.	A self-supporting student will be more aware of the work-income relationship.	-
<i>A</i> Older students will be less egalitarian than younger students.	Age is widely believed to be conducive to more conservative stances on political, economic, and social issues.	-
<i>O</i> Students whose parents are business operators or professionals will be less egalitarian.	Students coming from a business or professional home environment will anticipate greater personal benefit from participation in the market economy.	+
<i>I</i> Students from high-income families will be less egalitarian than students from low-income families.	Students who have benefited from their parents' relatively successful participation in the market economy will be reluctant to alter that system's income distribution.	-

TABLE 3
Attitude Survey Regression

Independent Variable	Coefficient Estimates	t Ratios
<i>S</i>	1.7470	2.4515*
<i>P</i>	0.8404	1.0538
<i>P</i>	-0.5657	-0.8099
<i>M</i>	2.6271	1.5395
<i>G_m</i>	-2.4819	-2.2481*
<i>G_h</i>	-2.6269	-2.1453*
<i>H</i>	0.0310	1.0846
<i>A</i>	-0.1889	-0.8563
<i>O_p</i>	1.7721	1.9988*
<i>O_w</i>	-0.0245	-0.0284
<i>I_m</i>	-1.3606	-1.3782
<i>I_h</i>	-2.1585	-2.0375*
Intercept	2.9408	0.6499
Standard error of estimate = 4.637		
$R^2 = .1233$; R^2 adjusted = .0642; $F = 2.086$		

*Significant at the .05 level.

G = cumulative grade average [two dummy variables: (a) $G_m = 1$ for medium grades in the 2.50-3.49 range on a 4.00 scale, $G_m = 0$ otherwise; (b) $G_h = 1$ for grades above 3.49, $G_h = 0$ otherwise; grades below 2.50 coded zero]

H = number of hours of employment per week

A = age of student in years

O = occupational status of parents [two dummy variables: (1) $O_p = 1$ if professional-technical, $O_p = 0$ otherwise; (2) $O_w = 1$ if "workers," defined as blue and white collar, farmers, etc., $O_w = 0$ otherwise; business parents coded zero]

I = student's perception of income status of parents [two dummy variables: (a) $I_m = 1$ if medium income, $I_m = 0$ otherwise; (b) $I_h = 1$ if high income, $I_h = 0$ otherwise; low income coded zero]

u = random error term

Table 2 states the ex ante hypotheses made for each independent variable, offers a brief rationale for each hypothesis, and indicates the anticipated coefficient sign if that hypothesis is confirmed by the data. There is some support in earlier research for the hypotheses associated with variables *S*, *G*, and *A* (Gifford et al.; Scott and Rothman). However, a recent study by Riddle (1978) suggests hypotheses which are the reverse of those stipulated for variables *P* and *M*.

The findings of the model for the experimental group are summarized in Table 3. We find that sex (*S*), cumulative grade average (G_m and G_h), occupational status of parents (O_p), and perceived income level of parents (I_h), all have the correct (hypothesized) signs and are statistically significant. The population (*P*), marital status (*M*), and age (*A*) variables have the correct signs,

but are not statistically significant. The race variable (R) also had the hypothesized sign, but the small number of nonwhites involved in the survey renders the statistical results highly suspect.³ The employment variable (H) is not significant and has the wrong sign.

The significant t values in the model indicate success in achieving the goal of identifying some very important influences on students' attitudes. The low R^2 's suggest that student values are subject to considerable random influences. Thus the model would be a poor predictor of attitudes.

Performance and Attitudinal Changes

If the difference in attitudes between the beginning and end of the semester is attributable in part to the economics course, it seems plausible that students who devoted more effort to the course (thereby receiving a higher grade) experienced the largest attitude shift. Furthermore, the level of students' attitudes at the semester's beginning may influence the size of the shift. An attempt was made to explain attitudinal changes ($DIFF$) by grade in economics (EG), as a measure of effort and understanding in the course, and preassessment scores (ATT).

Because incoming attitudes and course grades are related to student personal characteristics, both along with $DIFF$ must be regarded as endogenous variables in an equation system with personal characteristics as predetermined variables. The personal characteristics in Table 3 plus students' class in school were used as predetermined variables in a two-stage least squares procedure to yield:

$$DIFF = 2.0387 + .4245 ATT - .7577 EG; \quad R^2 = .0069$$

The t ratio for the coefficient on ATT was 1.9897, significant at the .05 level; that on EG was -2.6076, significant at the .01 level.

A lower (larger negative) $DIFF$ indicates a larger egalitarian attitude change. A disappointing R^2 indicates a large random element in $DIFF$ and probable imperfections in measurement. Also, it should be remembered that the two-stage least squares procedure, in contrast to ordinary least squares, does not maximize R^2 .

The significant coefficient results support the hypotheses. First, students with the least egalitarian incoming attitudes experienced the largest changes. Second, those with the highest cognition, measured by grades, also show the largest value changes.

Summary

This study suggests that (1) one semester of introductory economics may shift student attitudes on income distribution in an egalitarian direction; (2) sex, cumulative grade average, occupational status of parents, and perceived income level of parents are statistically significant determinants of student attitudes regarding income distribution; and (3) students' attitudinal changes varied directly with their performance in the course.

FOOTNOTES

1. See Okun (chap. 4). Most of one 50-minute lecture was devoted to these aspects of income distribution.
2. The 191 students represented 31 percent of the 601 students who took the ma

semester of the principles course in the fall of 1978. We have no reason to believe that these students did not constitute a random sample of the larger group.

3. Only five of the 191 students in the experimental group were nonwhites. Of these, only one was black and the remaining four were Asian-Oriental students. Also, only eight of 191 students were married. The students' ages ranged from 17 to 28, with the average age 19.7 years.

Our model initially contained a dummy variable distinguishing between business and nonbusiness students. This variable was dropped, however, because it is debatable whether curriculum choice should be treated as an independent or a dependent variable. Does a student develop an anti-egalitarian view as a consequence of enrolling in a business college or does his anti-egalitarian position prompt him to select that particular curriculum?

REFERENCES

- Gifford, John B., Robert L. Thornton, and H. Ralph Jones. "A Comparison of the Attitudes of Practitioners and Business Students Toward Social and Economic Systems," *Collegiate News and Views* 32 (Fall 1978), 1-9.
- McConnell, Campbell R. *Economics*. 7th ed. New York: McGraw-Hill, 1978.
- Okun, Arthur M. *Equality and Efficiency: The Big Tradeoff*. Washington, D.C.: Brookings, 1975.
- Riddle, Terry. "Student Opinions on Economic Issues: The Effects of an Introductory Economics Course," *J. Econ. Ed.* 9 (Spring 1978).
- Scott, James H., and Mitchell R. Rothman. "The Effect of an Introductory Economics Course on Student Political Attitudes," *J. Econ. Ed.* 7 (Spring 1975).

APPENDIX: ATTITUDINAL SURVEY

1. A nation's income should be distributed to its citizens in proportion to their contributions to the production of that output.
2. If a dollar is taken from a rich person and given to a poor person, the overall well-being of society will be increased.
3. The redistribution of income is a legitimate role for government in the context of the U.S. economy.
4. Poverty is a social disgrace in a wealthy society and should be eradicated through appropriate public policies.
5. The nation's annual income should be shared equally by all families comprising the economy.
6. Workers who make a current contribution to the national output should not be obligated to share the fruits of their productive effort with those who are not working.
7. The United States tax system should be reformed so that high-income families pay a smaller share, and low-income families pay a larger share, of total taxes than is now the case.
8. A negative income tax—wherein government subsidizes families whose incomes fall below a specified level—should be legislated by the Federal government.
9. A nation's income should be distributed to its citizens in accordance with their needs.
10. Families in the United States should not be guaranteed a minimum annual income.
11. Income should be distributed according to economic factors, not by political decisions.

An Analysis of the Marginal Products of the One- and Two-Semester Introductory Economics Courses

Ralph D. Elliott, M. Edwin Ireland, and
Teresa S. Cannon

Background Information and Introduction

Many colleges and universities offer two types of introductory courses in economics. These generally consist of the traditional two-semester sequence and a one-semester course. The two-semester course supposedly provides the student with the necessary technical tools for more advanced work in the field while the one-semester course usually highlights major economic issues and introduces economic concepts by developing alternative solutions to those real-world problems.

The purpose of this study is to determine whether students learn significantly more basic economic concepts in the traditional two-semester sequence than in the increasingly popular one-semester course. If not, the efficiency of our production process (teaching economics) can be improved and significant savings of student and faculty time and material costs can be achieved by substituting the one-semester for the two-semester course.

The results of past studies conflict. For example, Klos and Trenton [4] found that a good one-semester course can be at least as effective in improving student learning as the two-semester course. In contrast, Paden and Moyer [8] found that a two-semester course resulted in about a 10 percent improvement in test scores compared to a one-semester course. In addition, Dawson and Bernstein [2] in a study of introductory economics courses in colleges and universities in New York State, found that students who had completed a two-semester course scored significantly better on the Test of Economic Understanding (TEU) than students who had completed a one-semester course.

The present research effort is different from the previously mentioned studies in two ways. First, and perhaps most important, one-semester courses are significantly different now than they were in the late 1960s and early 1970s, when the other studies were carried out. Instead of being fast-paced, abbreviated versions of the two-semester sequence, many one-semester courses, including those at Clemson, are topic- or issue-oriented and no longer emphasize the mechanical and computational skills which have characterized the two-semester sequence. Furthermore, the one-semester course now has many books written exclusively for it. Instructors of such courses are no longer bound to select chapters from the traditional encyclopedia-type economics textbooks.

Another difference between the present study and previously published studies is our use of the Test of Understanding in College Economics (TUCE) as the means of evaluating student

Ralph D. Elliott is associate professor of economics and Teresa S. Cannon is a graduate student at Clemson University. M. Edwin Ireland is economist and manager of corporate economics at United Energy Resources, Inc.

performance, instead of the older Test of Economic Understanding (TEU). It is true that use of TUCE introduces many problems, the most important being that it was written specifically for the two-semester sequence. This no doubt biases downward the scores for the one-semester course. Other problems with TUCE arose from the nature of many of the questions themselves. Some were so ambiguous that many of our faculty could not agree on the correct answer. In some cases, the questions related to subjects that were not covered in the one-semester course. For example, international trade, which is weighted rather heavily in the macroeconomics part of the test, was given only a few lecture hours in our one-semester course.

In the study reported here, answers to the following specific questions are provided: (1) Are there observable differences in the level of basic economic knowledge gained in the two-semester course versus the one-semester course? (2) If so, is there a difference in (a) simple recognition of basic economic concepts, (b) simple application of economic concepts, (c) complex application of economic concepts?

Experimental Design

A total of 1,057 students were included in the experiment, with 169 selected from the one-semester course (Economics 200) and 888 from the two-semester courses. Of the latter, 677 were in the macroeconomics course (Economics 201). There were a total of 27 class sections involved in this study, 5 from Economics 200, 17 from Economics 201, and 5 from Economics 202. Leftwich and Sharp [5] and Rogers [9] were the two texts used in the one-semester course. Mansfield [6] and Miller [7] were the two texts used in the two-semester sequence.¹

At the beginning of the semester, students in these courses were given the appropriate TUCE pretests in order to provide data on this initial knowledge of basic economic concepts. Midway through the semester the students in the one-semester course were given the microeconomics posttest followed by the macro pretest, since the remaining part of that course is allocated to macroeconomic concepts.² At the end of the semester, all students were given the appropriate posttests. As a result of this testing, 311 complete scores on microeconomics exams and 470 complete scores on macro exams were collected.³

In order to delineate the factors influencing test scores, linear multiple regression analysis was used to match individual test scores with individual student characteristics, thus avoiding problems of aggregate analysis. The specified four forms of the dependent variable were (1) posttest scores; (2) absolute improvement ($ABS\ IMP = \text{posttest less pretest scores}$); (3) percent improvement [$\text{percent IMP} = (\text{posttest less pretest})/\text{pretest}$]; and gap closed = $(\text{posttest less pretest})/(33 \text{ less pretest})$. An analysis by Gery [3] found the gap closed variable to be the most reliable in that it eliminates the problems of convergence and heteroscedasticity that characterize the other three. In addition, it controls for the ceiling effect.⁴ Since we tried all four specifications, all results are presented for comparative purposes.

The explanatory or independent variables used in the analysis fall into three general categories: (1) human capital variables, (2) environmental variables, and (3) effort variables. Human capital variables were academic class, marital status, sex, college major, SAT scores, and pretest scores. Environmental variables were a course variable, high school size, class time, and class size. One effort variable was used—grade point ratio.

Since our primary goal was to test the impact on TUCE scores of the two-semester course versus the one-semester course, we have not hypothesized any particular relationship between the various independent and dependent variables. Instead we have concentrated on including all the various determinants of students' learning found in the literature on the model and have allowed the data to determine the signs on these variables.

Table 1
Macro Regressions

	Posttest	ABS Imp	Percent Imp	Gap Closed
Intercept	3.6953 (2.1909)**	3.6953 (2.1909)**	0.1120 (2.1909)**	0.0154 (0.1946)
Sec (1 sem)	-1.6473 (-5.559)***	-1.6473 (-5.559)***	-0.0500 (-5.559)***	-0.0868 (-6.225)***
Pretest Score	0.1721 (2.8587)***	-0.828 (-13.75)***	-0.025 (-13.75)***	-0.027 (-9.462)***
SAT scores	0.0832 (5.8318)***	0.0832 (5.8318)***	0.0025 (5.8318)***	0.0038 (5.6652)***
Grade point ratio	0.0084 (2.5198)***	0.0084 (2.5198)***	0.0003 (2.5198)***	0.0004 (2.6907)***
Class size ^a	0.0129 (2.0961)**	0.0129 (2.0961)**	0.0004 (2.0961)**	0.0007 (2.3135)**
High school size	-0.001 (-1.374)	-0.001 (-1.374)	-0.00004 (-1.374)	-0.00006 (-1.453)
Agriculture	0.5007 (1.5007)	0.5007 (1.0753)	0.0152 (1.0753)	0.0219 (0.9985)
Architecture	0.2187 (0.2440)	0.2187 (0.2440)	0.0066 (0.2440)	0.0084 (0.1999)
Education	-0.032 (-0.029)	-0.032 (-0.029)	-0.001 (-0.029)	-0.002 (-0.031)
Engineering	-0.102 (-0.200)	-0.102 (-0.200)	-0.03 (-0.200)	-0.009 (-0.369)
Industrial management and textile science	-0.589 (-1.586)	-0.589 (-1.586)	-0.018 (-1.586)	-0.023 (-1.797)
Liberal arts	-0.636 (-1.165)	-0.636 (-1.165)	-0.019 (-1.165)	-0.036 (-1.413)
Science	0.1969 (0.3743)	0.1969 (0.3743)	0.0060 (0.3743)	0.0124 (0.5017)
Freshman	-0.895 (-2.172)**	-0.895 (-2.172)**	-0.027 (-2.172)**	-0.043 (-2.210)**
Sophomore	-0.182 (-0.515)	-0.182 (-0.515)	-0.006 (-0.515)	-0.007 (-0.411)
Junior	0.2873 (0.6309)	0.2873 (0.6309)	0.0087 (0.6309)	0.0162 (0.7539)
Sex (male)	0.7122 (3.3194)***	0.7122 (3.3194)***	0.0216 (3.3194)***	0.0317 (3.1448)***
Marital status (single)	-0.524 (-0.626)	-0.524 (-0.626)	-0.016 (-0.626)	-0.024 (-0.608)
R^2	.2688	.3123	.3797	.3129

NOTE: Statistical significance: * = .10 level; ** = .05 level; *** = .01 level. Figures in parentheses are *t* statistics.

^aSince time and size were highly correlated, time was eliminated from the regression using stepwise procedures that maximized R^2 .

Empirical Results

Overall Ordinary Least Squares Regressions

The results of the regressions are summarized in Tables 1 and 2. In all cases R^2 is significant at the .01 level. In terms of significance, all models are consistent for both sets of data (macro and micro). The section variable, which tests for a difference between the two courses, shows significance at the .01 level for all equations.⁵ The coefficient is negative,

Table 2
Micro Regressions

	Posttest	ABS Imp	Percent Imp	Gap Closed
Intercept	5.5020 (2.9743)***	5.5020 (2.9743)***	0.1667 (2.9743)***	0.6531 (4.5003)***
Sec (1 sem)	-1.253 (-3.584)***	-1.253 (-3.584)***	-0.038 (-3.584)***	-0.653 (-2.795)***
Pretest score	0.1129 (2.0299)**	-0.887 (-15.96)***	-0.027 (-15.96)***	-0.074 (-17.04)***
SAT scores	0.0518 (3.5708)***	0.0518 (3.5708)***	0.0016 (3.5708)***	0.0039 (3.4107)***
Grade point ratio	0.0038 (1.0628)	0.0038 (1.0628)	0.0001 (1.0628)	-0.0001 (-0.573)
Class size	-0.002 (-0.325)	-0.002 (-0.325)	-0.0001 (-0.325)	-0.0004 (-1.003)
High school size	0.0008 (0.8302)	0.0008 (0.8302)	0.00003 (0.8302)	0.00004 (0.5278)
Agriculture	-1.412 (-0.880)	-1.412 (-0.880)	-0.043 (-0.880)	-0.067 (-0.536)
Architecture	-0.329 (-0.395)	-0.329 (-0.395)	-0.010 (-0.395)	-0.033 (-0.501)
Education	-1.688 (-0.846)	-1.688 (-0.846)	-0.051 (-0.846)	-0.105 (-0.668)
Engineering	2.1984 (3.2338)***	2.1984 (3.2338)***	0.0666 (3.2338)***	0.1578 (2.9593)***
Industrial management and textile science	0.2125 (0.3735)	0.2125 (0.3735)	0.0064 (0.3735)	0.0327 (0.7734)
Liberal arts	-0.341 (-0.458)	-0.341 (-0.458)	-0.010 (-0.458)	-0.053 (-0.902)
Science	1.0628 (1.5022)	1.0628 (1.5022)	0.0322 (1.5022)	0.0485 (0.8739)
Freshman	-0.990 (-1.334)	-0.990 (-1.334)	-0.030 (-1.334)	-0.100 (-1.716)*
Sophomore	-0.491 (-0.709)	-0.491 (-0.709)	-0.015 (-0.709)	-0.036 (-0.666)
Junior	-0.551 (-0.740)	-0.551 (-0.740)	-0.017 (-0.740)	-0.011 (-0.187)
Senior	0.3203 (0.3337)	0.3203 (0.3337)	0.0097 (0.3337)	-0.0267 (-0.354)
Sex (male)	0.3342 (1.4320)	0.3342 (1.4320)	0.0101 (1.4320)	0.0347 (1.8943)*
Marital status (single)	-0.368 (-0.398)	-0.368 (-0.398)	-0.011 (-0.398)	-0.17 (-0.233)
R ²	.3162	.5342	.5342	.5757

NOTE: Statistical significance: * = .10 level; ** = .05 level; *** = .01 level. Figures in parentheses are *t* statistics.

signifying that the two-semester students performed better on the TUCE test than did the one-semester students. In the case of the macro regressions, the gap closed coefficient is .09, while in the case of the micro regressions, it is .65. Although both coefficients are statistically significant, it is questionable whether the macro coefficient has any practical significance. In other words, the marginal benefits of an additional semester of macroeconomics are relatively small compared to the gain in microeconomics. Of course, both gains have to be compared to the marginal cost of an additional semester of work.⁶

Table 3
Macro Category Regressions, by Category of Questions

	POSTRU	POSTSA	POSTCA
Intercept	0.2203 (3.2260)***	0.0475 (0.6848)	0.0904 (1.2470)
Sec (1 sem)	-0.0265 (-2.252)**	-0.0618 (-5.033)***	-0.0551 (-4.333)***
Pretest score	0.0890 (1.6925)*	0.0641 (1.2244)	0.0841 (1.6112)*
SAT scores	0.0027 (4.7723)***	0.0027 (4.5999)***	0.0029 (4.7375)***
Grade point ratio	0.0002 (1.3489)	0.0004 (3.1556)***	0.0002 (1.4786)
Class size	0.0006 (2.5163)***	0.0002 (0.7036)	0.0004 (1.4500)
High school size	-0.001 (-1.701)*	-0.00005 (-1.241)	-0.00001 (-0.362)
Agriculture	0.0169 (0.9010)	0.0154 (0.7952)	0.0166 (0.8207)
Architecture	-0.008 (-0.217)	0.0142 (0.3799)	0.0120 (0.3038)
Education	0.0561 (1.2807)	-0.029 (-0.640)	-0.029 (-0.603)
Engineering	0.0044 (0.2154)	-0.001 (-0.040)	-0.081 (-0.365)
Industrial management and textile science	-0.025 (-1.680)*	-0.001 (-0.052)	-0.026 (-1.620)*
Liberal arts	-0.009 (-0.409)	-0.041 (-1.818)*	-0.010 (-0.430)
Science	-0.030 (-1.406)	0.0064 (0.2934)	0.0352 (1.5400)
Freshman	-0.010 (-0.620)	-0.053 (-3.058)***	-0.018 (-0.990)
Sophomore	0.0005 (0.0346)	-0.010 (-0.694)	-0.008 (-0.522)
Junior	0.0049 (0.2701)	0.0159 (0.8395)	0.0078 (0.3940)
Sex (male)	0.0218 (2.5406)***	0.0240 (2.8051)***	0.0242 (2.6066)***
Marital status	0.020 (-0.592)	-0.007 (-0.198)	-0.028 (-0.780)
R ²	.1427	.2090	.1645

NOTE: Statistical significance: * = .10 level; ** = .05 level; *** = .01 level. Figures in parentheses are *t* statistics.

Ordinary Least Squares Regressions by Category of Question

The category regressions are designed to demonstrate what influence the two-semester-versus-one-semester dummy variable as well as the other independent variables have on the posttest TUCE score when it is divided into the categories of difficulty i.e., recognition and understanding (RU), simple application (SA), and complex application (CA).⁷ The macro and micro results are shown in Tables 3 and 4. On all three question types, the macro results indicate that students taking a full semester of macroeconomics outperform the students taking about half a semester of the topic. The micro results indicate a significant improvement in

Table 4
Micro Category Regressions, by Category of Question

	POSTRU	POSTSA	POSTCA
Intercept	0.2164 (2.3429)**	0.2114 (2.3488)**	0.1292 (1.3900)
Sec (1 sem)	-0.038 (-2.173)**	-0.037 (2.093)**	0.009 (0.515)
Pretest score	-0.003 (-0.042)	0.112 (1.840)*	-0.033 (-0.514)
SAT scores	0.0012 (1.6415)**	0.0010 (2.1941)**	0.0026 (3.5641)***
Grade point ratio	0.0005 (2.6541)***	0.0003 (1.7406)*	0.0004 (1.9649)**
Class size	0.0005 (1.9568)**	0.0002 (0.8426)	0.0006 (2.1157)**
High school size	-0.0001 (-1.025)	0.0004 (0.8717)	-0.00001 (-0.260)
Agriculture	0.1062 (1.3232)	-0.020 (-0.254)	-0.044 (-0.537)
Architecture	-0.063 (-1.527)	-0.050 (-1.228)	-0.025 (-0.062)
Education	-0.064 (-0.645)	-0.141 (-1.441)	0.0548 (0.5377)
Engineering	-0.001 (-0.026)	0.0250 (0.7459)	-0.015 (-0.417)
Industrial management and textile science	-0.014 (-0.507)	0.0265 (0.9561)	0.0245 (0.8400)
Liberal arts	0.0434 (1.1714)	0.0483 (1.3278)	0.0385 (1.0173)
Science	-0.019 (-0.537)	0.0514 (1.4834)	-0.027 (-0.744)
Freshman	-0.002 (-0.042)	0.0008 (0.0222)	-0.055 (-1.460)
Sophomore	-0.019 (-0.555)	-0.014 (-0.408)	-0.031 (-0.864)
Junior	-0.003 (-0.071)	-0.015 (-0.405)	-0.046 (-1.222)
Senior	0.0004 (0.00091)	-0.043 (-0.924)	0.0136 (0.2780)
Sex (male)	0.0128 (1.0980)	0.0286 (2.1520)**	-0.0005 (-0.045)
Marital status (single)	0.0208 (0.4499)	0.0286 (-0.873)	0.0148 (0.3125)
R ²	.1681	.1856	.1534

NOTE: Statistical significance: * = .10 level; ** = .05 level; *** = .01 level. Figures in parentheses are *t* statistics.

performance on questions involving RU and SA questions. However, the coefficient on the section dummy for complex micro applications is not statistically significant. The result implies that one-semester students have the same grasp of complex micro concepts as do two-semester students.

The coefficients of determination are very low for these regressions. This would imply that limited weight should be given to the results except where they are consistent with previous findings.

Conclusion

With the exception of the micro equation for complex applications, all the section variables in the study were statistically significant. This means that students taking two semesters of micro- or macroeconomics may demonstrate a significant statistical improvement on their TUCE performance. However, students have to incur a 100 percent increase in cost—they must take a second semester of economics—in order to achieve the gains in performance found in the study.

Until the marginal cost and marginal benefits are quantified, only a value judgment can be made as to whether these gains are worth their added cost. However, these results imply that in some cases a reallocation of resources from the two-semester sequence to the one-semester course might prove beneficial. That is, if only the one-semester course were offered, then the concentration of teaching resources on it might improve the quality of that course sufficiently to offset its minor deficiency.

FOOTNOTES

¹The one-semester course is issue oriented: First, a problem is outlined as it is perceived by the public. The economic aspects of the problem are then discussed, after which a few basic economic tools are developed and applied to it. Micro concepts tend to receive more emphasis than macro concepts since over half of the issues discussed are micro oriented. In addition, the traditional income-expenditure model is not used to present macro concepts. Instead, an aggregate supply-demand framework is developed. The two-semester sequence is quite traditional and involves a discussion of basic concepts with some emphasis on application. In all classes, i.e., 200, 201, and 202, the traditional lecture method was used, with a single faculty member teaching the class throughout the term.

²Given that the one-semester-course students take microeconomics in the first part of the course, we would expect that they would perform better on the macro pretest than students who had never taken economics. (Six questions on the macro-TUCE are on microeconomics.) Hence, we might expect a smaller improvement on the macro test for those one-semester students than for two-semester students.

³As indicated above, 380 students were given the micro pretest and 846 were given the macro pretest. But only 311 micro and 470 macro complete observation sets were obtained for final analysis. The loss of observations occurred partly because some students were absent from the posttest or had dropped the course. A third reason was lack of complete background data on the student; in that case, the observation was deleted. A review of the backgrounds of the students who dropped the courses indicated there was no apparent difference in type between these students and those who did not drop the courses.

⁴A ceiling effect may exist for the other three dependent variables in that there are a finite number of questions on the TUCE, with the result that the higher the pretest score, the less room there is for possible improvement on the posttest. The gap closed dependent variable represents actual improvement divided by the potential improvement, thus indicating in percentage terms the extent to which the student closed the gap between the pretest score and the perfect score. See [1] for additional discussion on this issue.

⁵As indicated by the tables, various coefficients are significant in addition to the section variable. A discussion of the implications of the significance of these other independent variables is not included here, but is available from the authors upon request.

⁶Additional two-stage least squares regressions were completed on the data. The results were consistent with the one-stage models. The explanatory power of the models was not increased, and the section variable retained its level of significance. Copies of the two-stage results are available from the authors upon request.

⁷Ideally the gap closed dependent variable should have been used to test categories. However, this was not possible since there were an unequal number of questions in the various categories of difficulty between the pre- and posttest. This situation could have been avoided if the same test had been used as a pre- and posttest.

REFERENCES

1. Becker, William, Jr., and others. "A Debate on Research Techniques in Economic Education." *Journal of Economic Education*, Fall 1976.

2. Dawson, George, and Irving Bernstein. "The Effectiveness of Introductory Courses in High Schools and Colleges." New York: New York University Center for Economic Education, 1967.
3. Gery, Frank W. "Is There a Ceiling Effect to the Test of Understanding in College Economics." In *Research Papers in Economic Education*, edited by Arthur L. Welsh. New York: Joint Council on Economic Education, 1976.
4. Klos, Joseph J., and R. W. Trenton. "One Semester or Two." *Journal of Economic Education*, Fall 1969.
5. Leftwich, Richard H., and Ansel M. Sharp. *Economics of Social Issues*. Dallas: Business Publications, Inc., 1974.
6. Mansfield, Edwin. *Principles of Macroeconomics*. New York: Norton, 1974.
7. Miller, Roger L. *Economics Today: The Micro View*. 2nd ed. San Francisco: Canfield Press, 1976.
8. Paden, Donald W., and M. Eugene Moyer. "Some Evidence on the Appropriate Length of the Principles of Economics Courses." *Journal of Economic Education*, Spring 1971.
9. Rogers, A. J. *Choice*. Englewood Cliffs, N.J.: Prentice-Hall, 1974.
10. Sullivan, J. L. "Multiple Indicators and Complex Causal Models." In *Causal Models in the Social Sciences*, edited by H. M. Blalock Jr. Chicago: Aldine, 1971.

Student-to-Student Tutoring in Economics

Allen C. Kelley and Caroline Swartz

While student-to-student tutoring has been used with some success in several disciplines, there have been relatively few evaluations of its impact in economics. This note reports on the impact of using student-to-student tutoring in freshman economics at Duke University. We conclude that this rather simple to implement teaching technique conveys notable benefits to those being tutored. Indeed, in the case being reported, student-to-student tutoring conveys positive impacts on student achievement in economics which are more than twice as large as either (a) the impact of having had high school economics, or (b) the impact of having had high school differential calculus.

During the fall semester of 1975 the TIPS (Teaching Information Processing System) was used as a teaching tool in freshman economics at Duke University. TIPS is a computer-managed instructional system which facilitates collecting periodic (usually weekly) information on student achievement via multiple-choice surveys, which are not usually employed to determine course grades. This information is used to prescribe for each student an individualized course of instruction through a series of student reports. TIPS has been reported on elsewhere and, while it was integral to the course under study, is relatively unimportant to the main results analyzed below.¹ Suffice to say that through using the TIPS system, weekly "surveys" (multiple-choice quizzes) were administered, and the results of these surveys were used to identify those students who were performing very well in the course and those who were having some difficulty.

Those performing well on the TIPS surveys were provided the option of either taking the forthcoming one-hour examination or of being exempted from the examination and instead tutoring students who were having difficulty with the course material. Two alternative tutoring formats were used. For the first two portions of the course (approximately 12 of the 15 weeks) the tutors were required (a) to attend a one-hour training session, (b) to hold a one- to two-hour group tutorial, usually consisting of two tutors and three to six tutees, and (c) to be available on a one-to-one basis as requested by the tutees. In the third segment of the course the requirements of tutors were essentially the same, except that the group session was replaced by one hour where the tutor was available to tutees in a designated place to answer questions in a one-to-one situation. In this case the entire class was invited to use the tutorial service. Virtually all of those invited as tutors accepted this option.²

Those having difficulty with economics were invited, but not required, to attend the first two scheduled tutorial sessions. Around 18 percent of the class was provided this option over

Allen C. Kelley is Professor and Chairman of Economics, and Caroline Swartz is a graduate student, Duke University. They are grateful for the detailed comments by Professor Rendigs Fels on an earlier version of this note, and to Ms. April Brazell, who assisted with the statistical analysis.

the course of the semester. Of the class, 11 percent elected to participate in the tutorials. We will compare the performance of those who attended the tutorials with those who were invited but did not attend.

Those who did not attend the tutorials were self-selected. Thus, we do not have a controlled experiment whereby we can confidently abstract from student attributes which may be excluded from our study, but which may have influenced student achievement in economics, as well as the impact of tutorials. However, for two reasons, we feel that the results presented below may be meaningful. First, on *a priori* grounds, the impact of self-selection works in both directions, with possible offsetting results. On the one hand, those who did not attend tutorials may have been disorganized, busy, or less motivated, or possess some attribute which may result in an understatement of the impact of tutorials. On the other hand, those who did not attend may have felt correctly that they knew the materials sufficiently well, and would benefit relatively little from the tutorials. Their absence would result in an overstatement of the measured impact of the tutorials.³ Second, a comparison of the background attributes of those who attended the tutorials with those who did not (see Table 1) indicates that the two groups are very similar. There are no statistically significant differences in individual background attributes. This result, together with the findings below that a large portion of the class performance is explained by these background attributes, leads us cautiously to accept the hypothesis that the two groups are basically similar, and that the results presented below reflect, without notable bias, the impact of the tutorial program.

An alternative, more skeptical (and possibly more realistic) view of the experimental design would hold that there is self-selection bias since student motivation, an important element in explaining student achievement, is not adequately measured by the attributes considered in Table 1. As a result, the findings presented below which show students in the experimental group performing higher than those in the "control" group should be reinterpreted to represent the combined impact of tutorials and the somewhat higher motivation by those students taking advantage of tutorials. Since it is the motivated students who are likely to benefit most from tutoring, or any teaching technique under evaluation, this experiment might reflect a relatively efficient format for offering the tutoring option. The results may therefore be interpreted in this context.

The course required students to take three examinations (two midterms and a final), totaling 120 points. Students were also required to hand in five cases; the scores on the best four (80 points maximum) were included in the total points for the course. The mean and standard deviation of the class total points for the course were 177.3 and 13.0, respectively.

Equation (1) presents an explanation of the student's total score (T) based on background attributes (listed in Table 1), and on whether or not the student attended one or more of the

$$\begin{aligned}
 (1) \quad T = & 117.7 + .04 \text{ SATVERB} + .03 \text{ SATMATH} + 5.7 \text{ HSGPA} + 3.98 \text{ MATH1} \\
 & (3.76) \quad (2.33) \quad (3.43) \quad (2.37) \\
 & + 8.66 \text{ MATH2} + 3.23 \text{ ECONB4} - 13.10 D_1 - 8.32 D_2 - 9.39 D_3 - 8.24 D_4 \\
 & (2.85) \quad (2.09) \quad (-2.95) \quad (-2.19) \quad (-2.79) \quad (-3.66) \\
 & + 2.51 D_5 \\
 & (.97)
 \end{aligned}
 \qquad R^2 = .52$$

tutorial sessions. The background attributes are all statistically significant with the expected sign. Having had economics or differential calculus in high school contributes 3.23 or 3.98 points, respectively; having had integral calculus contributes an additional 4.68 points. The impact of the tutorial program, totaling 8.44 points, is measured as the sum of difference in the estimated parameters for D_1 and D_2 , D_3 and D_4 , and D_5 .⁴ D_1 and D_3 equal unity for those students who were invited and did not attend the first and second tutorials, respectively; D_2 and

Table 1
Background Attributes of the Class, the Tutored Group, and the Nontutored Group

Background Attribute	Mean Value of Attribute, Standard Deviation in Parentheses			
	Class (1)	Tutored* Group (2)	Nontutored* Group (3)	Difference Between (2) and (3)?**
SAT verbal score (SATVERB)	590.3 (75.8)	545.0 (93.3)	535.4 (85.3)	No
SAT quantitative score (SATMATH)	644.5 (76.3)	593.2 (74.1)	603.1 (95.3)	No
High school grade point average (HSGPA)	3.41 (.45)	3.17 (.47)	3.16 (.47)	No
Highest math attainment: differential calculus (MATH 1)	.30 (.46)	.14 (.36)	.23 (.44)	No
Highest math attainment: integral calculus (MATH 2)	.06 (.24)	.04 (.19)	0 (0)	No
High school economics (ECONB4)	.30 (.46)	.39 (.50)	.39 (.51)	No

*The "Nontutored Group" refers to those who were invited to attend either of the first two tutorial sessions but who did not attend any of the three sessions. Invitations to attend only the first two are used because these invitations were selective of students having difficulties based on TIPS survey results, whereas the entire class was invited to the third session.

**Is the difference in the indicated attribute statistically significant at the 95 percent level?

D_1 equal unity for those students who were invited and attended the first and second tutorials, respectively. D_3 equals unity for those students who attended the third tutorial. (The entire class, excepting the appointed tutors, was invited to participate as tutees.)

The estimated impact of the tutorial program is large. It is equal to 4.2 percent of the possible total points for the course, and to two-thirds of the standard deviation around the actual mean total point score for the class. It also has a greater impact than having had both high school economics and high school differential calculus. Moreover, its impact is larger, as measured by the percentage increment to the student's course score, than the impact of alternative modifications in the technology of teaching as reported in many studies in this *Journal*. Finally, in terms of the student's grade in the course, the tutorial program in many cases would have raised the student's formal evaluation by a full letter grade.⁵

The results presented above are likely to understate the impact of the tutorials since they do not reveal the effect of the program on the tutors. Professors John J. Siegfried and Stephen H. Strand, in an evaluation using data from Professor Rendigs Fels' course at Vanderbilt University, have discovered that it is the tutors who benefit the most from the program.⁶ Indeed, in their experiments, the tutees did not benefit from this technique.

We have little to say about the optimal format for a tutorial program. Ours is but one experiment with what we assert is a reasonably low cost format, and one which is likely to yield relatively high returns. Hopefully, over time, sufficient information will be available from other experiments to form a reasonable judgment on its impact and best format in the teaching of economics.

Footnotes

¹Allen C. Kelley, "TIPS and Technical Change in Classroom Instruction," *American Economic Review*, 62 (May 1972), 422-428; Allen C. Kelley, "Individualizing Instruction through the Use of Technology in Higher Education," *Journal of Economic Education*, 4 (Spring 1973), 77-89.

²This is somewhat surprising since these students were told that the total time they would spend tutoring would notably exceed the amount of time that they would likely spend studying and taking the midterm examinations. Moreover, the grade the tutors would receive would not be a top score, but would vary within the "A" range, depending on the evaluation of their sessions by the tutees.

³This analysis assumes that the disorganized, less motivated students who stayed away would have benefited more than the average of those who attended, and the bright students who stayed away would have benefited relatively less than those who attended.

⁴In an exploratory regression model we also attempted to ascertain whether the students being tutored benefited differently according to whether or not their high school GPA score was relatively high or low. Unfortunately, the relatively low variance of GPA scores (and several other background attribute scores) for the group being tutored constrained obtaining a reliable evaluation of the possible existence of interaction effects. D_3 is statistically different from D_1 at the 95 percent level; D_4 is not statistically different from D_3 .

⁵Several other "output" measures were investigated (e.g., scores on each examination, total scores on examinations, total score on cases), and the results are basically the same as the summary result presented in equation 1. For example, a regression was run using only the scores on the first two exams as the dependent variable and omitting the D_3 variable. The results are:

$$\begin{aligned} \text{EXAMS} = & 34.84 + .02 \text{ SATVERB} + .02 \text{ SATMATH} + 3.12 \text{ HSGPA} + 2.40 \text{ MATH1} \\ & (3.06) \quad (2.88) \quad (3.25) \quad (2.47) \\ & + 4.64 \text{ MATH2} + 1.87 \text{ ECONB4} - 8.40 D_1 - 4.54 D_2 - 4.01 D_3 - 4.19 D_4 \\ & (2.64) \quad (2.09) \quad (3.26) \quad (2.09) \quad (2.05) \quad (3.23) \end{aligned}$$

$$R^2 = .50$$

where EXAMS = sum of first two examinations (80 points maximum). This model has the benefit of confining the evaluation to the case where the group invited to the tutorials was selected solely on the basis of TIPS survey results. The average increment to the student's potential score was 4.6 percent ($3.7/80 \times 100$), a figure consistent with the results presented in the text. It should be noted that the second examination in the course was exceptionally difficult.

⁶John J. Siegfried and Stephen H. Strand, "An Evaluation of the Vanderbilt-JCEE Experimental PSI Course in Elementary Economics," see this Journal, pages 9-26. The impact of the program on the tutors will be published in a forthcoming issue of the *Southern Economic Journal*.

Textbooks and the Teaching of Economic Principles

Marian R. Meinkoth

The economic principles course in the large university frequently exhibits a high degree of standardization. A common text or texts, a common topical outline, a common reading list, and a uniform examination may be used. Such standardization tends to suppress experimentation by individual instructors and, as the number of students and the size of the teaching staff expands, experimentation on a course-wide basis becomes very difficult to engineer. Many object to such standardization as too rigid. At Temple University the requirement of a uniform textbook for the Principles of Economics course has been considered particularly irksome by many members of the teaching staff. On the other hand, permitting each instructor freedom to choose his own text and other reading materials has been viewed with some apprehension, particularly since advanced undergraduate courses in economics depend upon the principles course to provide basic groundwork.

This paper reports on an experiment devised to demonstrate the effect of using various textbooks chosen by instructors. In the experimental group were five faculty members teaching the first half of the principles course. While they varied in rank from instructor to full professor, all had had several semesters of experience teaching in principles. During the first semester each used Samuelson [1]. During the second semester, one continued to use Samuelson although on a more selective basis than before, two used Heilbroner [2], one used Alchian and Allen [3], and one used three paperbacks by Mundell [4], Schultz [5], and Gill [6].

Part I, forms A and B, of the Test of Understanding in College Economics (TUCE) [7] was used at the end of the course. Each participating faculty member was given a list of the categories covered in the TUCE¹. For the first semester all instructors used the

Marian R. Meinkoth is Associate Professor of Economics at Temple University.

¹They are: scarcity, functioning of economic systems, basic elements in supply and demand; macroeconomic accounting; determination of GNP (income-expenditure theory); money, banking and monetary policy; government fiscal policies; determinants of economic growth; and policies of stabilization and growth.

uniform one-page outline of chapter assignments for Samuelson, covering chapters 1-19 and 37-40. But for the second semester each was left completely free to determine the sequence of coverage, to allocate the time to be devoted to each topic, and to make other reading assignments. None saw the final exam before it was given in either semester, so there was no possibility of bias arising because of familiarity with the examination.²

Table 1 compares the mean and standard deviation at the end of each semester for each individual professor, and for the total experimental group, as well as the TUCE norms. The results are surprisingly similar for the two semesters.³ For Part A the mean for the entire group tested was 20.05 for both the first semester and second semester. For Part B the result was 20.74 for the first semester and 20.11 for the second. Means for individual instructors showed somewhat more variation but did not change by statistically significant amounts from the first semester, except in one instance. In the case of professor B₁, who was replaced by professor B₂ in the second semester, there was a statistically significant difference at the .05 level in the mean score on Part A of the exam, but not on Part B.

The author is familiar with only one other study in which the use of different textbooks was compared [8, p. 220]. That study, however, focused primarily on a comparison of the results of self-study using programmed learning texts versus a conventional textbook in a lecture or discussion course. The conventional textbooks included three "well-known" texts, designated as A, B and C, used by 60 percent of the students, and ten textbooks designated as "all other," used by the other 40 percent. The conclusion was that the textbook used appeared to have significant effects on TUCE scores. The mean score for all students taking the test was 17.9 out of a maximum possible score of 30. Students using text "A" scored .86 point higher, and those using "C" .75 point higher than the average of students using texts in the "all other" category—a significant difference—while students using text "B" scored only .42 higher, which is not a significant difference. However, differences between textbook "B" and textbooks "A" and "C" were not significant. Since these results were apparently not adjusted for other variables, the type of school and quality of students using a particular textbook could well have influenced the mean score of students using the book in question.

A questionnaire distributed to the students during the final examination at the end of both semesters indicated a slightly more favorable attitude toward the textbook used during the second semester. Those students who indicated that the textbook was "helpful in understanding the main ideas of the course" increased from 58 to 74 percent. Those who found the material in the text difficult to understand dropped from 40 to 36 percent, and those who gave the textbook an overall satisfactory rating increased from 57 to 63 percent. The slightly more favorable attitude toward the texts used in the second semester may reflect the fact that an instructor utilizes the text of his choice more efficiently than one which is chosen by someone else.

On the other hand, the change in textbook policy does not seem to have influenced the attitude of the student toward economics. Relative to other current courses, economics was rated first or second in difficulty by 56.3 percent of all students in the fall, and 57.9 percent in the spring semester. Experience with the course influenced 14.7 percent in the fall semester and 14.5 percent in the spring semester to "take a voluntary

²Graduate students were used as proctors the first time the examination was given.

³A detailed comparison of the personal characteristics of the two experimental groups of students will be mailed to anyone requesting it if he will send a stamped self-addressed envelop to Prof. Meinkoth, Department of Economics, Temple University, Philadelphia, Pennsylvania 19122.

Table 1
Academic Year 1968-69

Professors†	Number of Students	Part A		Part B	
		Mean	SD	Mean	SD
First Semester Exam					
A	54	19.33	4.19	18.96	4.32
B ₁	37	21.35	3.93	20.34	3.99
C	104	20.43	4.78	21.13	4.30
D	103	19.26	4.99	20.89	4.70
E	30	21.10	4.77	22.58	4.79
Total	328	20.05	4.71	20.74	4.52
National norm A	876	19.16	5.39	— —	— —
National norm B	829	— —	— —	19.41	5.30
Second Semester Exam					
A	71	18.18	4.12	18.69	4.17
B ₂	38	19.42	4.19	19.84	4.31
C	114	20.12	4.49	20.68	4.15
D	61	19.84	4.75	21.24	4.84
E	31	22.35	5.40	21.36	5.24
Total	315	20.05	4.64	20.11	4.47
National norm A	876	19.16	5.39	— —	— —
National norm B	829	— —	— —	19.41	5.30

†Professor B was unable to continue with the experiment the second semester, so the first-semester participant is designated as B₁, and the second-semester participant as B₂.

The textbooks used by individual professors during the second semester of the year were: Professor A, selected material from [1]; Professor B₂, [3]; Professors C and D, [2]; and Professor E, [4-6].

elective in economics," caused 49.1 percent in the fall and 48.2 percent in the spring to "read more economic material in newspapers and magazines," but "persuaded" only 1.3 percent in the fall and .9 percent in the spring to major in economics.

Clearly, the relaxation in choice of text caused no great change in either student attitude or test results. There is also no evidence that permitting the individual instructor to choose his text has a detrimental effect on the course. If a uniform text is not used in the principles course, however, it seems desirable to agree on the broad topics to be covered in each semester. In addition, a list of optional topics for each semester would permit the instructor some individual choice without duplication between the two semesters. If a department feels that some checks are necessary in order to assure a degree of uniformity in the coverage of basic economic concepts, it seems preferable to test the end result rather than try to control the way in which that end is achieved.

This suggests a common uniform examination. Hopefully, there will be additional updated versions of the TUCE to be used for this purpose. It has the advantage that all questions have been carefully tested in advance, and, as long as individual instructors have not seen the test in advance, of giving no advantage to the students of any one instructor. But it is useful only when its coverage coincides with the basic material that a department wishes its students to master. The alternative is an examination prepared by the department, which tends to favor the students of those who play the largest role in its preparation. In either case, the exam used to test the achievement of minimum basic, and hence hopefully common, knowledge should constitute only a portion of the student's

course grade, which should reflect not only the results of the common examination but all other judgments of the individual professor as to a student's achievement.

Conclusion

The results of this experiment suggest:

(1) that there may be a wide range of reading materials which can be used successfully for teaching the basic macro- and microeconomic theory included in the principles course in economics;

(2) that, given as a guide a list of broad topics such as the categories included in the TUCE, there will be little or no significant difference between the achievement (as measured by a similar uniform examination) of students taught from one standard textbook and those using a variety of texts and/or other materials chosen by their individual instructors;

(3) that permitting the instructor to choose his own teaching materials permits him to experiment more freely than if he is committed to a text chosen by someone else;

(4) that the students likewise may receive benefits from the more creative and innovative approach permitted the instructor; and

(5) that using some form of uniform examination at the end of the course can be used to achieve the minimum amount of uniformity that is desired.

REFERENCES

1. Paul Samuelson, *Economics: An Introductory Analysis*, 7th Edition. New York: McGraw-Hill, Inc., 1967.
2. Robert L. Heilbroner, *The Economic Problem*. Homewood, Ill.: Richard D. Irwin, Inc., 1968.
3. Armen A. Alchian and William R. Allen, *University Economics*. Belmont, Cal.: Wadsworth Publishing Co., Inc., 1967.
4. Robert A. Mundell, *Man and Economics*. New York: McGraw-Hill Book Co., Inc., 1968.
5. Charles L. Schultz, *National Income Analysis*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1967.
6. Richard T. Gill, *Economics and the Public Interest*. Pacific Palisades, Cal.: Good-year Publishing Co., Inc., 1968.
7. *Test of Understanding in College Economics*. New York: The Psychological Corporation, 1968.
8. Richard E. Attiyeh, G.L. Bach and Keith G. Lumsden, "The Efficiency of Programmed Learning in Teaching Economics: The Results of a Nationwide Experiment," *AER*, 59 (May 1969), 217-223.

Teaching Principles of Economics: The Joint Council Experimental Economics Course Project

By ALLEN C. KELLY*

For the past several years the Joint Council on Economic Education has engaged in a project to identify and assess alternative approaches to the teaching of college introductory economics. The goals of the project, as summarized by Arthur Welsh, have been "... to develop alternative approaches that overburdened professors in two- and four-year colleges might find more useful than their current offerings and to encourage others to improve and expand upon the Joint Council's efforts" (Rendigs Fels, p. 1).

Several professors and schools have participated in this effort: Kenneth and Elsie Boulding of the University of Colorado, Rendigs Fels of Vanderbilt University, Richard H. Leftwich and Ansel M. Sharp of Oklahoma State University, Phillip Saunders of Indiana University, and Barbara and Howard Tuckman of Florida State University. Syllabi and supporting materials have now been published as special issues of *The Journal of Economic Education* for the last four of these courses (Fels, Leftwich and Sharp, Saunders, Tuckman).¹ The course developed by the Bouldings was reported upon in preliminary form at the December 1973 American Economic Association meetings, and thus I will concentrate my attention in this review on the remaining four.

I. Overview and Comments on the Individual Course Packages

There is a common set of premises underlying the four courses which have been devel-

oped. Fels summarizes these premises well. "Standard textbooks are typically overloaded. All too often instructors feel obliged to assign the whole book, leading to overloaded courses. As a result, the student gains vague familiarity with a wide range of economic theory and a mastery of none of it. In addition, the typical course provides no training in the skills of applying economic principles" (Fels, p. 5).

The course developed by Fels carries this theme to its most extensive level. This course is structured around teaching the *application* of economic theory to a wide range of realistic problems. Students review "cases" which take the form of relatively short expositions of economic situations—often based on quotations from newspapers, or on edited newspaper articles. Students then work through a carefully constructed set of questions designed to train them in the process of orderly thinking about economic problems. To make the teaching approach more easily adopted by others, a casebook and instructor's manual have been developed, coauthored by Fels and Robert G. Uhler. Fels's syllabus shows in detail how cases may be incorporated into a "conventional" course (i.e., the traditional lecture format), identifies the key concepts which should be drawn out for each lesson, and even provides sets of notes to aid professors in leading class meetings.

In addition to the case application emphasis, Fels has developed his course around the Personalized System of Instruction (PSI) approach pioneered by Fred S. Keller. PSI employs virtually no lectures. The lecture period (if used for formal instruction) can be devoted to other activities: discussion, test taking, tutoring, projects, and so forth. Students may proceed at their own pace. Their performance is evaluated on the number of course units mastered. Assessment of mastery learning over the various in-

*Professor and Chairman, Dept. of Economics, Duke University. I am grateful for the comments of W. Lee Hansen, John J. Siegfried, and Sue Whitesell on an earlier draft of this paper.

¹To obtain a copy of the four syllabi, write Publications Department, Joint Council on Economic Education, 1212 Avenue of the Americas, New York, NY 10036.

structional units is accomplished by written or oral examinations. Students have considerable flexibility in selecting the time to take the examinations; these may also be retaken any number of times, until mastery achievement has been demonstrated. Because of the extensive amount of testing required in such a course, student undergraduate proctors are used. The proctors are also available for student consultation and tutoring. They receive three credits toward their economics major for participating as proctors. Budget constraints will typically preclude using exclusively higher-paid graduate students in this role. There is one proctor for approximately ten students. Ten proctors is around the maximum a professor can supervise; thus, *PSI* classes do not usually exceed one hundred students.

We are fortunate to have available, even at this early state, an evaluation of Fels's course by John J. Siegfried and Stephen H. Strand, neither of whom was involved in the course development or in its implementation (Siegfried and Siegfried and Strand). Their careful studies reveal that

"... (1) *PSI* students performed no better or no worse on multiple choice or essay examinations than students in the conventional lecture course; (2) there was no difference in performance in subsequent economics courses...; (3) students liked the course more, thought they learned more, and felt they were examined and graded fairer in the *PSI* course; (4) there was no difference in the amount of time spent on the course activities between *PSI* and conventional course students; (5) the tendency to elect economics as a major was unrelated to the method of instruction...; and (6) the student-proctors learned more economic theory than they would have learned from an alternative upper class economics elective." [Siegfried, pp. 32-33]

The latter effect, while based on a small sample, is quantitatively quite large.

Fels's contributions lie in two quite separate areas: the application of *PSI* to economics, and the development of the case-application approach. In a sense he has provided two course packages.

Of the four course packages under review, Fels's course is not only the most innovative, but also the most complete and diffusible. It includes the course syllabus, illustrative exami-

nations, notes to professors to guide their discussion sessions, cases, instructions to students, and so forth.

A few relatively minor qualifications might be expressed relating to Fels's course. First, the course is highly labor intensive—especially the *PSI* format, but to a lesser extent the case approach. To obtain proficiency in economic analysis, students must write several cases; these must be graded, and students must be provided detailed feedback. While I am convinced that the case-applications approach represents a superior way of teaching the most important elements of economic principles, I would still like more information on alternative formats of using this approach which employ less labor inputs. I suspect that a totally case-oriented course with active student involvement, and a course which is also economically feasible in a wide range of colleges and institutions, is yet to be discovered. I hope that considerable experimentation with alternative formats will be stimulated by the high quality materials made available by Fels and Uhler. Equally important, I hope that professors evaluate the effectiveness of these alternative formats, and then report their results.

A second qualm relates to the difficulty of the materials. While Fels has provided suggestions for lowering the difficulty level, I doubt if these suggestions will be sufficient. He has set his standards high. More experimentation will be required to develop a less demanding set of course objectives—possibly by trimming some of the content, but not necessarily the level of required analysis. Parenthetically, some would seriously question the feasibility of the goal of teaching freshmen and sophomores how to engage in simple and complex application of economic principles. My own position is that this goal may not be feasible for the majority of this group of students. However, the value of the benefits obtained for those who achieve this level of proficiency will far exceed any foregone learning of "tools" usually taught in the principles course, since these tools, typically taught without extensive application, are rapidly forgotten for lack of purpose.

Finally, it must be kept in mind that the *PSI* approach, while promising, is costly. Fels has

noted that course instructors should probably be awarded more than one course "credit" to provide them the incentive and time to implement *PSI*. Moreover, critical to this approach is the availability of student proctors. It is extremely encouraging that Siegfried's research has shown that the benefit of proctoring to the learning of economics notably exceeds the opportunity cost of the typical upper division economics course. More hard evidence on this point from other studies could be decisive in making college administrators and faculty receptive to providing course credit for proctoring, and thus effectively opening up the financial and technological viability of *PSI*.

The course by Leftwich and Sharp centers around the "issues approach" to teaching economics, and also stresses application of economic principles. They have developed a book which focuses on the various issues taken up in their syllabus. They have also provided references to several other books which offer pertinent cases and materials. Their issues are generally broader than Fels's cases, and as a result, there are fewer of these issues, and each one occupies around a week of the course. The syllabus itself provides for each issue "major discussion points," "economic concepts and principles" covered, and a "recapitulation." While these aids are useful, the course package would have been considerably more valuable and adaptable by others had the authors also provided guidelines and suggestions to the professor for actually organizing and leading the various discussion sessions, together with test items which evaluate the students' mastery of the various issues. The issues, however, are well chosen and should engender considerable student interest. Moreover, in unpublished research results, the authors have argued that the exclusive teaching of economics around these several issues does not result in any notable sacrifice of the basic tools typically taught in the standard economics course.

Saunders' course confronts the difficult problem of developing and coordinating an application-oriented offering involving the participation of as many as twenty different instructors. He has assembled with the participation of his colleagues a consensus on a required "core" of

economic analysis which is common to all sections, and which is tested by a common final examination. Each professor is then free to develop specific application emphases. e.g., environmental economics, current economic issues, income distribution. Students are free to select the orientation that interests them most. A review of the alternatives open to students at Indiana University reveals a smorgasbord of courses seldom found in even the combined offerings of a dozen institutions of higher education.

Saunders' contributions lie mainly in his interesting and useful presentation of information on course planning, teaching techniques, and research formulation. He has presented valuable tips, in highly readable form, on such topics as the formulation and evaluation of examinations, the development and assessment of course evaluations, the preparation of course objectives, the construction of research designs for appraising teaching, the techniques of coordinating courses with many professors, the alternative ways to use and train graduate student instructors, and so forth. Some useful base-line data are also presented on course evaluation surveys and examinations.³ In my judgment, his syllabus is most helpful on the general methods of course development, and especially on the techniques of coordinating and influencing many instructors to adopt a common core of economic analysis, yet allowing each professor to "do his or her own thing." This format makes faculty more amenable to undertaking the perceived "chore" of teaching principles of economics, and at the same time, increases their effectiveness.³

³For example, on a point that has received considerable attention (*J. Econ. Ed.*, Fall 1973), Saunders' data collected over eight semesters from almost nine thousand students indicate that student performance on the common final examinations is positively and significantly associated with the evaluation rating of the instructor from whom they took the course.

³Subsequent to the publication of the syllabus being reviewed here, Saunders has developed two student workbooks—one for microeconomics and one for macroeconomics—that define the analytical core for each semester of Indiana's introductory course in terms of specific examination questions and a set of 15–20 homework problems. A special seminar has also been developed to train graduate student instructors to become more effective teachers of introductory economics.

Barbara and Howard Tuckman's course is possibly the most traditional of the ones offered, and as a result, will be quite easily adopted by many institutions. Its innovations are less in the course content area, and more in the areas of providing techniques for motivating students to learn economics, and providing students flexibility in their pace of studying economics and of taking examinations. Students are placed in a simulated environment of being actual consultants to the government and are required to complete "memoranda" (disguised workbook and/or case application-like exercises of a policy orientation) for the President and his key advisors. The authors also offer a few specific and useful ideas for arousing student interest during the classroom presentation of some key economic tools: the consumption function, the multiplier, and so forth. Their attempts at self-pacing of student learning met with mixed results. When provided freedom of time allocation students procrastinated, creating administrative and testing burdens at the conclusion of the course. The authors have thus elected to offer rewards and penalties (in the form of course points) which have the net effect of bringing many of the students back toward the traditional study mode in terms of time allocation. Their findings accord with my own hunch that the self-paced mode is suitable primarily for the more highly motivated and disciplined students—and these, lamentably, are not the vast majority of students in American higher education.

II. Overall Appraisal of the Joint Council Program

With the exception of Fels's package, each of the remaining syllabi is somewhat incomplete in providing all the elements of a "turn-key" offering for the overburdened economics instructor. However, I do not consider this a telling deficiency of the developmental effort. Something should be left for the instructor to do, overburdened or not. More important, I find that the various packages, *when taken together*, provide a surprisingly wide variety of materials and ideas that a large number of instructors should find useful in their course planning and

implementation. For example, the instructor will find very helpful Saunders' and Tuckmans' techniques of course planning and course coordination with many teachers; the instructor may utilize some of Fels and Uhler's cases, and some of the issues provided by Leftwich and Sharp; and the instructor will also hopefully engage in some course evaluation of the type illustrated by Siegfried, Strand, and Saunders.

Other conspicuous achievements of the Joint Council program are both the respectability it has provided the idea of teaching more limited course content with higher emphasis on application, and its many suggestions on ways for accomplishing these goals. Two books containing useful case and issue materials have resulted from the project; others are now available from commercial publishers, in part in response to the project itself. While the optimal format for using the case approach may not yet have been discovered, this approach is here to stay, and will gradually improve over time.

Finally, these syllabi will serve to stretch the thinking of the teacher trapped in the routine of providing students with a comprehensive coverage of encyclopedic texts. Such instructors, with the aid of these Joint Council materials, can rediscover some of the excitement of teaching. For example, while the Tuckmans' Presidential Policy Memoranda do not represent a notable breakthrough in concept coverage, this technique does indeed illustrate an innovative way of "packaging" economics to engender some increased enthusiasm and interest—and that contribution is not to be discounted. These syllabi will encourage other professors to break out of their traditional mold, and to discover new techniques of teaching economics in a more exciting and effective manner. Certainly the syllabi demonstrate conclusively that this field is wide open, and that this activity can be intellectually rewarding.

In pointing to the future of the experimental courses under review, and to other developmental efforts of this type, two broad issues should be raised. First, it would be good to establish a modest program to track over time the successes and failures of these courses. One might, for example, document over a period of

five years the experience of a group of adopters, asking each to respond periodically to a carefully constructed series of questions. Several issues come to mind. How many of the original set of adopters continue to use the materials after five years? Why do they continue (or stop) using the materials? What role do the adopters play in changing the original course packages, and in what specific ways?

Second, a review of these syllabi raises anew the critical question of how best to evaluate and compare various courses and teaching approaches. While instruments such as the Test of Understanding in College Economics are useful, this test is not by itself sufficient to compare such divergent courses as those represented by the syllabi under review. The time is ripe in economic education research 1) to identify a fairly comprehensive list of outputs which are likely to be impacted by changes in course content and teaching technique, 2) to develop reliable and standardized instruments and procedures for measuring these outputs, and 3) to attempt to obtain some weights on the value of the outputs as provided by the various clientele of our teaching programs. Such a "Manual of Instructional Outputs and Their Measurement" would include items ranging from student enrollments, the number of majors, student "enjoyment" of the course, faculty willingness to teach the course, to such items as student problem solving skills in general, knowledge of economic tools, ability to apply economic tools, social and political values, and the like.⁴ The ability to utilize the wide range of research results that have been forthcoming in the *Journal of Economic Education* and elsewhere is increasingly being constrained by the heteroge-

neity of the set of instructional outputs chosen by various researchers to evaluate their teaching, and the wide range of techniques employed to measure even quite similar outputs. Comparisons are therefore notably hampered. We should begin thinking about *systematizing* our efforts in identifying, measuring and weighting the various outputs of education in general, and of economic education in particular. A careful evaluation of the Joint Council experimental courses could represent an excellent application of, and stimulus to, such an outcome.

REFERENCES

- Kenneth and Elise Boulding**, "Introducing Freshmen to the Social System," *Amer. Econ. Rev. Proc.*, May 1974, 64, 414-19.
- Rendigs Fels**, "The Vanderbilt-JCEE Experimental Course in Elementary Economics," *J. Econ. Ed.*, Winter 1974, Special Issue No. 2.
- and **Robert G. Uhler**, eds., *Casebook of Economic Problems and Policies: Practice in Thinking*, St. Paul 1976.
- Richard H. Leftwich and Ansel M. Sharp**, *Economics of Social Issues*, Dallas, 1974.
- and —, "Syllabus for an 'Issues Approach' to Teaching Economic Principles," *J. Econ. Ed.*, Winter 1974, Special Issue No. 1.
- Phillip Saunders**, "Experimental Course Development in Introductory Economics at Indiana University," *J. Econ. Ed.*, Fall 1975, Special Issue No. 4.
- John J. Siegfried**, "Is Teaching the Best Way to Learn? An Evaluation of the Benefits and Costs to Undergraduate Student Proctors in Elementary Economics," *Southern Econ. J.*, January 1977.
- and **Stephen H. Strand**, "An Evaluation of the Vanderbilt-JCEE Experimental PSI Course in Elementary Economics," *J. Econ. Ed.*, Fall 1976.
- Barbara and Howard Tuckman**, "Toward a More Effective Principles Class: The Florida State University Experience," *J. Econ. Ed.*, Spring 1975, Special Issue No. 3.

⁴Attention to developing and systematizing such a set of instruments will be provided in a recently undertaken five-year project founded by the National Institute of Education, and undertaken by Richard Attiyeh, Keith Lumsden, and myself. We will be developing several alternative course packages for teaching economics, employing Teaching Information Processing System, programmed learning, computer games, cases, and the conventional lecture technique. Several dozen schools will be participating in the testing and evaluation of these various packages. The ability to identify, measure, and compare instructional outputs across course packages is clearly critical to the research.

Test Information:

An Application of the Economics of Search

William F. Barnes

Economic analysis is applied in explaining increasing varieties of human behavior—for example, marriage, fertility and crime. Frequently the applications identify the gains and costs, and utilize the marginal framework to discuss resource allocations. Recently the marginal framework has been applied to explain an individual's behavior in acquiring additional information when confronted by imperfect information [5]. The approach has been widely used in discussing search behavior in the labor market [7, 9]. This framework appears to have a useful application in the study of student search for, and acquisition of, information about test questions. In this paper the acquisition of test information is analyzed by identifying the gains and costs associated with investment in search. The consideration of search gains and costs predicts explanatory variables of the acquisition of information. Implications are tested against empirical evidence from the performance of college students on tests in economics classes.

Numerous psychological studies explain cheating among college students as deviant behavior [1, 3, 4, 6, 10]. In most studies a questionnaire sample survey of self-reported behavior provides the empirical base. The association of cheating to personality characteristics is emphasized. The relation of cheating to demographic variables is reported less frequently, and with little explanation. Statistical tests are typically two-way cross tabulations. In contrast, this paper treats the acquisition of test information, which is only one form of cheating, as rational, though not admirable, behavior. Data are from an experiment that permits cheating behavior to be observed independently rather than self-reported as in a questionnaire survey. Predictions are tested by a multivariate statistical technique.

Identifying Test Information Acquirers

Empirical evidence was obtained from written examinations of college students in six sections of a one-quarter (10-week) junior-senior level course in labor economics, which was required with a C or better for most students enrolled in the business program, but not required for students recently entering the college because of a curriculum change. The six sections were offered two sections per quarter for three quarters. One instructor presented the same lectures in all sections. Class size varied from 48 to 131 students.

For the two sections of the same course offered at different times of the day 50 percent of the multiple-choice questions were included in both the exams given. The remaining questions were different for each of the two exams. Question sheets were collected when students turned in answer sheets. Each quarter all questions were new. Students who acquired test information about the test (rather than knowledge of the subject) were identified by their performance

William F. Barnes is Assistant Professor of Economics, The University of Georgia.

relative to the classes' mean performance on exam questions for which information was potentially available, compared to their performance (relative to the mean performance) on questions for which information was not available.¹ This procedure was followed in the first test of the quarter for three quarters.

Test information was potentially available to students in the second section of the day from students in the first section. In the second section each student's score for the repeated questions (y) was compared to the section's mean score ($\bar{y}$), calculated by the ratio $y/\bar{y}$. Also, each student's score for the new questions (x) was compared to the section's mean ($\bar{x}$), calculated by the ratio $x/\bar{x}$. The ratio $(y/\bar{y})/(x/\bar{x})$ shows the student's performance relative to the class on the repeated compared to the new questions. The expected value of the ratio is 1, when the student's relative score on the new questions is the same as that on repeated questions. Larger ratios show better relative performance on the repeated than on new questions. Those students in the second class of the day who have ratios significantly greater than by chance are presumed to have acquired information about the test from students taking the test earlier. Since the test information acquired by some raises the class mean on repeated questions ($\bar{y}$), the expected value of the ratio for a noncheater is lowered. Two approaches in identifying cheaters are possible. The first is to find ratios significantly greater than the mean by accepting the upward biased estimate for $\bar{y}$. This would underestimate the number of cheaters. The second is to use as the estimate of the class mean score on repeated questions ($\bar{y}$) the mean of the first class. The second procedure was selected because it can be assumed that since both classes are samples from the same population, the estimate $\bar{y}$ would not be biased. Using $\bar{y}$ in the first class as an estimate for $\bar{y}$ in the second class, the expected value of $(y/\bar{y})/(x/\bar{x})$ for noncheaters in the second class is 1. To determine whether persons have ratios significantly greater than 1 (the expected mean value for noncheaters), an estimate of the variance is necessary. The larger ratios of cheating students in the second class will raise the class variance. The variance of the first class provides an unbiased estimate of the variance in ratios of the noncheaters, which can be used to compute the confidence interval to determine whether the individual ratios of the second class are significantly greater than 1 (the mean for noncheaters).

This ratio (with estimated $\bar{y}$ of first class) was calculated for each student in the second class. Students performing relatively better on the repeated than on the new questions (whose ratios were significantly greater than one at the .01 level using the estimated variance of noncheaters) were classified as cheaters.² The number and percentage of cheaters together with the class size are reported below:

	<i>Second Class Size*</i>	<i>Number of Cheaters</i>	<i>Percentage of Cheaters</i>
First quarter	81	18	22.2
Second quarter	49	19	38.8
Third quarter	131	57	43.5
Total	261	94	36.0

*Sizes of the first class of the day were 83, 79 and 48.

Theory of Test Information Acquisition

Thirty-six percent of students enrolled in the classes for whom this type of cheating was possible did cheat. The extent shows that such information gathering is typical, not deviant, behavior. This suggests the acquisition of test information can be analyzed using the standard theory of search behavior.

Time spent acquiring information to improve exam performance is presumed to increase the value of schooling. The maximizing strategy for a student is to allocate resources so that the

marginal cost of acquiring test information is equal to the marginal expected return from the information. Time allocated to improve exam performance may be used to cheat as well as to study. Both activities may be viewed as investments in schooling. Ethical questions aside, time could be allocated to cheating and to studying, subject to the least cost condition that marginal physical products divided by factor prices for cheating and studying be equated. Returns to acquiring information generally and to cheating specifically may be almost impossible to estimate. Nevertheless, factors may be identified which affect the (rate of) return, the productivity and cost of cheating compared to studying and thereby increase or decrease cheating.

Factors that would raise the return from exam performance and increase the student search for test information include the requirement of the course for a degree and being nearer to graduation. Part of the return from schooling may depend upon the degree itself, completion of the program. The return from exam performance and therefore the search for test information may be greater in a required course since passing is necessary to obtain the college degree. As the student nears graduation the return from exam performance may be greater because of returns from the degree itself which depend on completion of the program. Students may also attempt to improve grades as graduation approaches, believing prospective employers will take recent performance as the more reliable index of future performance.

Reliance on cheating is also affected by the relative productivities and costs of acquiring test information compared to studying. Students with higher grade point averages, who are taking the course for the first time, or who are majors should be relatively more productive in studying compared to cheating than are students with lower grade points, who are repeating the course, or who are nonmajors. Higher grade point averages indicate ability to study. The inadequate previous performance of course repeaters suggests low productivity in studying. Majors have more experience in subject matter and probably some preference to learn course material. Also, cheating may have a lower return for majors who may expect to use the course material during later courses. It is not so likely nonmajors will encounter course material in later courses. The cost of this form of cheating compared to studying is lower for students with more friends in the first section and for students whose opportunity cost of time is greater. Since studying is expected to require more time than acquisition of test information, working students' higher opportunity cost of time may encourage cheating. Friends in the earlier class reduce the search time required to gain access to test information. The length of time prior to the second section exam limits the acquisition of test information. Also a longer period of time should reduce the cost to acquire test information. With less time between classes, acquiring test information may require additional costs due to missed classes. More time also provides more opportunity for test information to be acquired in casual encounters.

Empirical Results

Data on these factors are provided by a questionnaire administered in each class, seeking information on explanatory variables for each student. This information provided for the construction of the proposed explanatory variables.

Students likely to acquire test information are those for whom the return to test performance is greater—the course is required (REQ) and graduation is nearer (GRAD)—and for whom cheating is more attractive than studying—working students (JOB), lower grade point students (GPA), course repeaters (REP), persons with a longer period of time between the first and second section of the day (TIM2 and TIM4), nonmajors (MAJ), persons with friends in the first class (FRI), and men (SEX).

Linear discriminant analysis may be used to determine whether hypothesized variables distinguish between students who acquired test information and those who did not. This technique classifies students into two groups, those who did and who did not acquire test information on the basis of the hypothesized characteristics. Discriminant analysis provides a

Table 1
Linear Discriminant Function for
Student Acquisition of Test Information

Variable	Coefficient	Variable	Coefficient
Constant	41.533	MAJ	-19.171
JOB	-12.531	SEX	-16.620
REQ	4.658	TIM2	3.982
GRAD	6.493	TIM4	7.014
GPA	-5.942	FRI	8.551
REP	25.606		

function which is linear in the characteristics and also provides a critical value of the function used for classifying a particular individual as acquiring or not acquiring test information. The form is estimated by regressing a dummy variable for acquisition of test information (a value of 1 signifies information was acquired and of 0 signifies no information was acquired) on the characteristics, which are treated as explanatory variables. A linear discriminant function was obtained for acquisition of test information and is reported in Table 1.³

In interpreting discriminant functions, a large positive coefficient on a variable indicates students with higher values of this variable have a greater probability of being classified as acquirers of test information, *ceteris paribus*. A large negative coefficient implies students with low values of this variable have a greater probability of being nonacquirers of test information, *ceteris paribus*. The direction of all impacts estimated by the coefficients in the function for acquisition of test information are, except for JOB, consistent with *a priori* hypotheses. The coefficients on REQ, GRAD, REP, FRI, TIM2 and TIM4 are positive, while the coefficients for GPA, MAJ and SEX are negative.

The coefficients can be interpreted as the impact of the respective variables on the probability of an individual having acquired test information. The function suggests the probability that an individual acquired information increases 4.66 percentage points in a required course and 6.49 percentage points for being a senior. Being a course repeater increases the probability 25.61 percentage points. Having friends in the earlier class raises the probability 8.55 points. In comparison to no intervening period a two-hour interval before the second section increases the probability of acquiring test information 3.98 percentage points; a four-hour period increases the probability 7.01 percentage points. The probability falls 5.94 percentage points for each increment of one point, including the first one, in the overall grade point average. Other factors which reduce the probability of acquiring test information are being a major, by 19.17 points; currently having a job, by 12.53 points; and being a female, by 16.62 points.

Two tests of significance are available for discriminant analysis. For the discriminant function an *F*-test shows whether the two populations, the test information acquirers and the

Table 2
Confusion Matrix

Actual Acquisition	Classified Acquisition		Percent Observations Correctly Classified
	Acquired	Did Not Acquire	
Acquired	71	23	75.5
Did not acquire	54	113	67.7
Percent of all observations correctly classified			70.5

nonacquirers, are significantly different. An F -value of 12.19 with 9 and 250 degrees of freedom indicates that the two groups are significantly different for acquisition of test information. Also, the value from the discriminant function serves to classify individuals into one of the two groups. The ability of a discriminant function to distinguish correctly between the two groups may be shown in a confusion matrix [8]. This indicates the number of individuals classified correctly and incorrectly as well as where misclassifications occurred. The matrix reported in Table 2 shows the performance of the discriminant function, which classified 70.5 percent of the students correctly. This compares favorably to the 53.9 percent chance criterion for being correctly classified.

Conclusions

Ninety-four students in a sample of 261 sought and acquired sufficient information about a test given earlier in the day to improve their scores significantly for the repeated questions. The effort put into the search was even greater than that measured since half the questions identified were not repeated and since some of the questions probably could have been answered from knowledge of the subject. The considerable search effort occurred despite the opinion of most students, 94.3 percent, that the examinations were fair.⁴ Students apparently sought test information not because they felt exams were unfair and not as an exception but rather as a normal part of preparing for an exam. This behavior raises several challenges.

The problem of cheating is not so great with motivated students—majors, students with higher grade points, those voluntarily in a class. The problem is more serious in required courses, with more repeaters and students looking toward graduation. Unfortunately it is in the former case with fewer sections and smaller class sizes that essay exams are most plausible. It is in the latter case, where the problem is more serious, that large class size necessitates objective exams. If the same questions must be used, sections should be scheduled with no intervening period.

This study measured only one type of cheating, acquiring test information from students who took a test earlier in the day. The extent of this type suggests other types of cheating may be more widespread than previously indicated in studies of self-reported behavior. There was little risk, as there is in some forms of cheating, in seeking information about an earlier test in the day or about tests from the preceding quarters. If test information is sought as aggressively from those earlier tests (which may even be available in complete and written form), considerable amounts of test information may be acquired. This makes the repeated use of test banks questionable, unless the banks are available to all.

If this behavior were victimless, acquisition of test information would not be a problem, but it is sufficiently widespread to call into question the equity of grades. If normal distributions are used in assigning grades, noncheaters will be penalized. A similar problem would occur if grades were assigned less formally on the basis of relative performance.

Finally it is particularly interesting that fewer students with jobs, despite the higher opportunity cost of their time and the reduced time available for study, acquired test information. Apparently they are less interested in a high score and more interested in learning (exam scores and working were positively correlated, $r^2 = .481$). The students with jobs who presumably are in a better position to evaluate the importance of formal schooling for the labor market, attach more importance to learning and less to higher scores achieved through acquiring test information. If the learning aspect of schooling is more valued in combination with working at a job, then the schooling experience may be more intensely utilized.

Footnotes

⁴Efforts were made to restrict other types of cheating. Students were separated and exams were carefully proctored. The study is limited in that acquisition of information about the test earlier in the day is the only form of cheating measured. This certainly is not the most seriously regarded form of cheating. Some may not call the behavior cheating, but most would consider it unfair.

²The concern of this paper is to identify and investigate students who cheat (acquire test information) rather than to see whether sufficient information was gathered by members of the second class to raise the class mean score significantly on the repeated questions or on the entire test. It is nevertheless interesting to note that for all three quarters, while there was no significant difference in scores between first and second sections on new questions, for the repeated questions the mean score for the second section was significantly higher (at the .01 level). The better average performance on the repeated questions indicates that cheating, the acquisition of test information, did occur, but indicates nothing about the identity and characteristics of cheaters.

³Discriminant analysis is discussed in [2] and [8]. The discriminant function sometimes is reported as the linear regression with coefficients multiplied by a constant. It is reported here as the linear regression so effects of explanatory variables can be interpreted as probabilities.

It is interesting to test the statistical significance of these coefficients in *t*-tests: All but two of the 10 coefficients are significant at the 0.05 level. This procedure is questionable, however, because dichotomous dependent variables violate the assumption of normality of error terms.

⁴This opinion was available from anonymous, classwide teacher evaluations administered each quarter.

REFERENCES

1. D.L. DeVries and I. Ajzen, "The Relationship of Attitudes and Normative Beliefs to Cheating in College," *Journal of Social Psychology*, 83 (April 1971), 199-207.
2. P. J. Dhrymes, *Econometrics*. Evanston: Harper and Row, 1970.
3. R. A. Dienstbier and P. O. Munter, "Cheating as a Function of the Labeling of Natural Arousal," *Journal of Personality and Social Psychology*, 17, No. 2 (1971), 208-213.
4. J. Harp and P. Taietz, "Academic Integrity and Social Structure: A Study of Cheating Among College Students," *Social Problems*, 13, No. 4 (1966), 365-373.
5. J. Hirshleifer, "Where Are We in the Theory of Information?" *American Economic Review*, 63 (May 1973), 31-39.
6. J. Q. Knowlton and L. A. Hamerlynck, "Perception of Deviant Behavior: A Study of Cheating," *Journal of Educational Psychology*, 58, No. 6 (1967), 379-385.
7. J. J. McCall, "Economics of Information and Job Search," *Quarterly Journal of Economics*, 84 (February 1970), 113-126.
8. D. G. Morrison, "On the Interpretation of Discriminant Analysis," *Journal of Marketing Research*, 6 (May 1969), 156-163.
9. S. C. Salop, "Systematic Job Search and Unemployment," *Review of Economic Studies*, 40 (April 1973), 191-201.
10. C. I. Stannard and W. J. Bowers, "The College Fraternity as an Opportunity Structure for Meeting Academic Demands," *Social Problems*, 17, No. 3 (1970), 371-390.

EDUCATIONAL PRODUCTION FUNCTIONS

Solomon W. Polachek Thomas J. Kniesner Henrick J. Harwood

University of North Carolina at Chapel Hill

Key words: Educational Production Function, Student Performance, Trade-offs in the Educational Process, CPES Production Function, Nonlinear Estimation

ABSTRACT

This research examines scholastic performance within the context of an individual's production function. A constant partial elasticity of substitution production function for academic achievement is presented and estimated with nonlinear maximum likelihood methods. We find that ability and time devoted to various aspects of the learning process are the most important determinants of students' accomplishments. Our results underscore the potential for students to compensate for relatively "poor" educational backgrounds by spending more time on study and class attendance.

INTRODUCTION

There exists in the social sciences a rich empirical and theoretical literature concerning the distribution of personal income. No matter what the approach, whether it be human capital (Mincer, 1974), statistical decomposition (Lydall, 1968), or eclectic socioeconomic analysis (Ginties, 1971), education surfaces as a prime quantifiable determinant of earnings differences within a population. This link between earning power and education underscores the importance of understanding the educational process.

Two basic views characterize the current state of economic analysis of education. First, education is treated as the output produced by a school. Second, the individual student is viewed as using his or her own time and effort, along with resources purchased from a school, to produce learning. Both approaches may be conceptualized by what economists call a production function. Economic theory places restrictions on the mathematical form of a production function, and we shall first examine briefly those restrictions. Next, we present estimates of educational production functions with data from a survey of undergraduate students at the University of North Carolina, Chapel Hill, corroborated by estimates from a national survey of college students.

Our research goals are to illuminate the value of the production function concept in helping to understand the educational process and to demonstrate that the production function concept provides some guidance in formulating public policy designed to influence the quality of education.

SCHOLASTIC PERFORMANCE: A PRODUCTION FUNCTION APPROACH

Consider an individual college student whose particular intelligence, amount of study time, and utilization of other resources leads to some level of academic success or scholastic performance. The mapping of this student's educational inputs (X) into a degree of scholastic performance (y) is a production function

$$y = F(X), \quad (1)$$

where $F(\cdot) \equiv$ the production function.

In general, there are two mathematical properties of a production function worth noting. First, it is single valued, continuous, and well defined over the input set, yielding nonnegative outputs. Second, it has continuous first and second-order partial derivatives. (See Ferguson, 1969 chapter 4) for a useful background reference on the technical properties of production functions.) Figure 1 depicts a typical production function in two inputs. A first partial derivative of $F(\cdot)$ is known as a marginal product because it indicates the productive effect of increasing that input, all other input levels held constant. Positive marginal products will typically be observed, as negative marginal products indicate wasted resources.

A key property of equation (1) is that it indicates that a particular level of performance may be produced in a variety

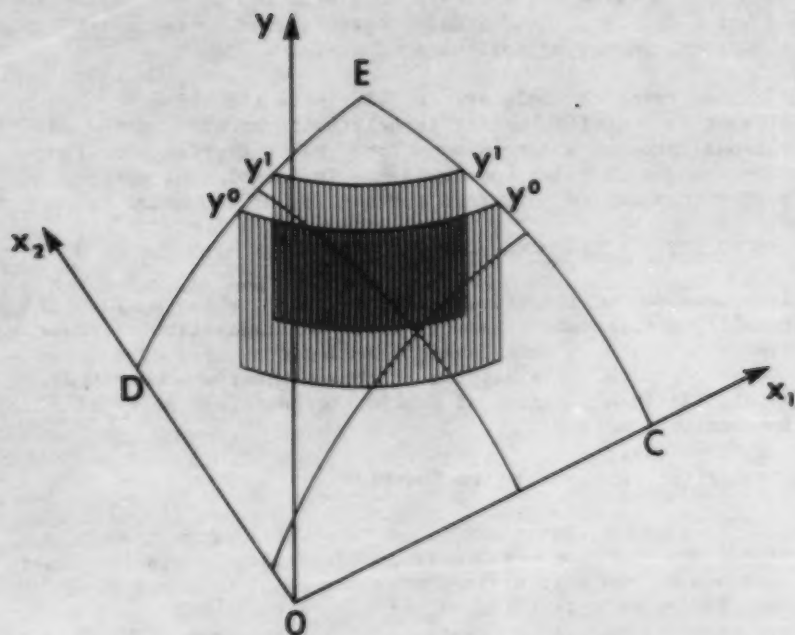


FIGURE 1

Production Surface ODEC Depicting Production Function for Output y with Two Inputs (x_1, x_2)

of ways. A grade of "B" may result from intensive home study coupled with sporadic class attendance or from perfect attendance paired with little study. What this means is that a student may trade-off class attendance for study time, or vice versa, in some proportion and still obtain a given grade. Such a trade-off is measured by a marginal rate of substitution (MRS). To see this characteristic of a production function, totally differentiate equation (1), set dy equal to zero, and solve for $(dX_j/dX_i)|_{\bar{y}, \bar{X}_{k \neq i,j}}$ yielding

$$\text{MRS} \equiv \left(\frac{dX_j}{dX_i} \right) \Big|_{\bar{y}, \bar{X}_{k \neq i,j}} = - \frac{\partial y / \partial X_i}{\partial y / \partial X_j} \quad (2)$$

Equation (2) gives us the increase in X_j necessary to hold y constant if X_i is decreased by a small amount. Figure 2 depicts two possible iso-output projections (isoquants), which are combinations of the inputs X_1 and X_2 that yield two particular levels of scholastic performance, y^0 and y^1 . The isoquant labeled y^0 , for example, is obtained from Figure 1 by projecting into X_1, X_2 space all (positive) combinations of the two inputs that lie along a locus of constant height (y^0). The slope of an isoquant at a particular point, say A in Figure 2, is a graphical representation of the marginal rate of substitution of X_1 for X_2 .

A standardized measure of the substitutability property is the elasticity of substitution (σ). This is defined as the percentage change in the ratio of two inputs resulting from a 1% change in the slope of the isoquant (marginal rate of substitution). In Figure 2, σ is the percent difference in the slope of ray OA and the slope of ray OB, divided by the percent difference in the slopes of isoquant (y^0) at A and at B. The elasticity of substitution basically reflects the curvature of an isoquant. The easier it is to substitute one input for another, the greater the value of σ is, and the closer to a downward sloping straight line the isoquant is. "Perfect" substitution is said to exist when σ goes to infinity. At the other end of the spectrum is a production function whose isoquants are right angles ($\sigma=0$), indicating that inputs must be used in fixed ratios if waste is to be avoided. In general, production functions are classified according to whether σ is constant or varies along an isoquant. In our empirical research to follow, we make use of two different constant partial elasticity of substitution (CPES) production functions: the Cobb-Douglas and the generalized CPES (Cobb

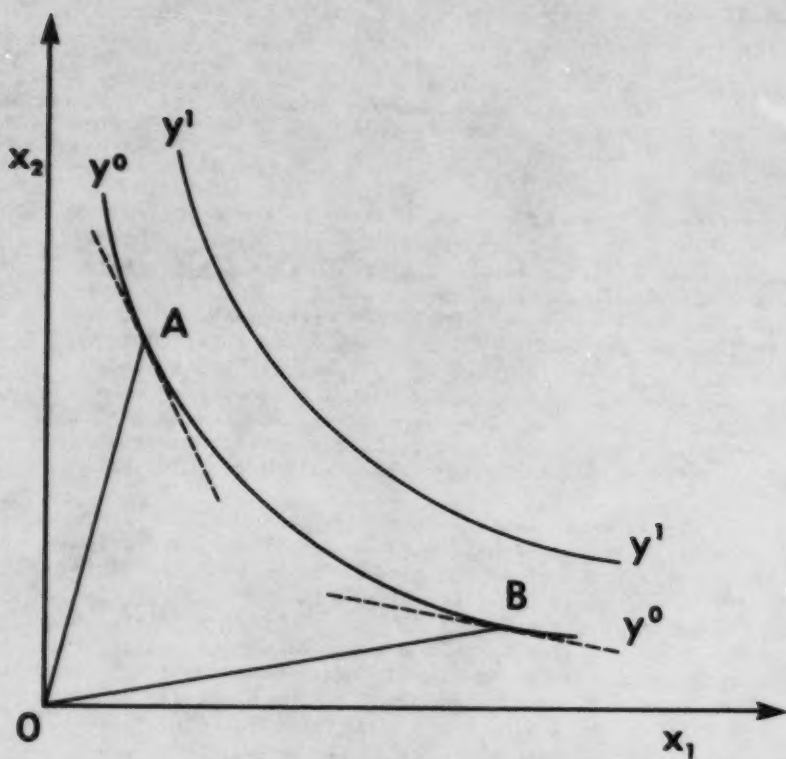


FIGURE 2

Iso-Output Projections of Input Combinations(x_1, x_2) Yielding Given Outputs y^0 and y^1

& Douglas, 1928). In the Cobb-Douglas function, σ is fixed at 1.0, a priori, and in the CPES the data are permitted to indicate an elasticity of substitution that takes on the same (constant) value for all pairs of inputs. Whereas the functions we employ are not the most general for the purpose of examining the intricacies of the educational process, they do provide the basic insights we seek, while remaining relatively easy to estimate.

THE CPES PRODUCTION FUNCTION

Equation (3) is a CPES production function for academic success. This formulation was suggested by Uzawa (1962); it was selected because empirical experimentation with more complex functions (see Ferguson, 1969, p. 110) proved unfruitful.

$$y = \gamma \left[\sum_{i=1}^n \delta_i X_i^{-\rho} \right]^{-\mu/\rho} \quad (3)$$

where $y \equiv$ academic success

$X_i \equiv$ inputs of personal attributes, scholastic environment, and study time

$\gamma, \rho, \mu, \delta_i \equiv$ (positive) parameters

$i = 1, \dots, n \equiv$ number of inputs.

This function has marginal products which are positive and diminishing:

$$\frac{\partial y}{\partial X_i} \equiv MP_i = \frac{\mu + \rho}{X_i^{\rho+1}} \mu \delta_i \gamma^{-\rho/\mu} \quad i = 1, \dots, n. \quad (4)$$

From the equations summarized in expression (4), the marginal rates of substitution of X_i for X_j are

$$MRS_{ij} = \frac{\delta_i}{\delta_j} \left(\frac{X_j}{X_i} \right)^{\frac{1}{\sigma_{ij}}} \quad i, j = 1, \dots, n, \quad i \neq j. \quad (5)$$

where $\sigma_{ij} \equiv \frac{1}{1+\rho}$.

From equation (5), it can be shown that σ_{ij} is the partial elasticity of substitution of X_i for X_j . To see this, differ-

entiate equation (5) and solve for $\frac{d \ln(X_j/X_i)}{d \ln(MRS_{ij})}$, which is the definition of the elasticity of substitution. If one takes the limit of equation (3) as ρ goes to zero, the Cobb-Douglas

production function $y = \gamma \prod_{i=1}^n X_i^{\delta_i}$ emerges. Production theory,

as summarized in equations (4) and (5), serves as a guide in our subsequent empirical analysis of the marginal productivities of a student's resources and the potential to compensate for certain background deficiencies.

EMPIRICAL IMPLEMENTATION: THE CPES

Overview

This section is devoted to estimation of equation (3). We apply nonlinear maximum likelihood techniques to a unique set of data containing detailed information on student personal attributes, time allocation, and scholastic performance. We utilize test scores on a standardized midterm examination in a large lecture course to measure scholastic performance. (See Bowles, 1970, for a survey of this issue.)¹ Because the academic achievement of individual students within a given class setting is analyzed, broad measures of school quality are implicitly held constant. So, we examine mainly the roles of ability and time across students.

One of the peculiarities of an empirical study of a microeconomic educational production function is the paucity of data. No information exists on how much time students devote to study and class attendance. Because of such data limitations, we created a unique body of information by surveying students in the principles of economics course at the University of North Carolina at Chapel Hill. In addition, a national data set (Eckland, 1972, and Eckland & M. J. Gillivray, 1972) that contains qualitative information is used to corroborate the empirical results from our survey.

At the University of North Carolina at Chapel Hill, the principles of economics courses are primarily taught with a lecture-seminar system. Students attend a large common

¹Good grades have been shown by Wise (1975) to exert some positive influence on lifetime earnings independent of their tendency to encourage more education and thereby affect earnings indirectly. By contributing positively to a higher course grade, the production of a "good" examination score is consistent with greater future economic welfare for students.

lecture presented twice a week by a faculty member and a small one-hour seminar given by graduate teaching assistants. We surveyed 227 students taking a particular macroeconomic principles lecture plus one of several associated seminar sections during the spring semester, 1975. The period of the survey covers the part of the course (five weeks) between the beginning and first hourly examination. By limiting the survey to such a short period of time, we have minimized, it is to be hoped, measurement errors stemming from respondents' ability to recall class attendance and study time.

Data concerning sex, test score, study time, lectures attended, college board scores, and socioeconomic background were gathered; summary statistics are displayed in Table I. It should be noted that in conducting the survey we employed a double blind procedure which guaranteed that students involved could not be identified by name. One side-effect of this was that test reliability and validity measures could not be constructed.

Statistical Methodology

We utilize as the empirical counterpart of equation (3)

$$G_t = \gamma \left[\delta_1 L_t^{-\rho} + \delta_2 S_t^{-\rho} + \delta_3 C_t^{-\rho} \right]^{-1/\rho} + \epsilon_t \quad (6)$$

$t = 1, \dots, T$

where $t \equiv$ index of observation and
 $\epsilon_t \equiv$ random error term.

The dependent variable (G_t) is the proportion of correct responses on a 50-question objective test. This score reflects the learning that occurred during the first five weeks of the course and is one component of the student's course grade. L is the number of hours the student spent in class (lectures plus seminars) during the portion of the course prior to the examination. Study time (S) represents the total number of hours a student studied specifically for the first examination. S may best be interpreted as "cramming" time. The last of the three independent variables, C , is the individual's score on the quantitative portion of the SAT test, which is required of all UNC-CH students. To keep the empirical analysis as simple as possible at first, we are parsimonious in specifying the inputs in the grade production process. Later on, we control

TABLE I
Summary Statistics of UNC-CH Data

	The Sample Pooled			The Sample Stratified					
				Males			Females		
	<u>Mean</u>	<u>Standard Deviation</u>	<u>Range</u>	<u>Mean</u>	<u>Standard Deviation</u>	<u>Range</u>	<u>Mean</u>	<u>Standard Deviation</u>	<u>Range</u>
G	84.90	8.66	43	85.80	8.33	43	83.75	9.00	42
L	12.52	1.85	12	12.67	2.11	12	12.33	1.43	9
S	5.35	3.50	19	5.05	3.54	19	5.74	3.43	19
C	579.50	81.49	430	587.10	80.06	410	569.69	82.68	500

where

G $\equiv$ numerical grade on midterm examination

L $\equiv$ lectures plus seminars attended

S $\equiv$ hours studied for the examination

C $\equiv$ score on quantitative section of Scholastic Aptitude Test

for sex and family background differences among students.

We assume that ϵ is independently and normally distributed with mean zero and constant variance σ^2 . Equation (6) is intrinsically nonlinear, and we choose to estimate its parameters by nonlinear least squares.² This technique finds the values of the six parameters in equation (6) that minimize

$$S(\underline{\theta}) = \sum_{t=1}^T \{G_t - f(\underline{X}_t, \underline{\theta})\}^2, \quad (7)$$

where $T \equiv$ the number of observations

$$\underline{\theta} \equiv [\gamma, \delta_1, \delta_2, \delta_3, \rho, u]'$$
 and

$$\underline{X}_t \equiv [C_t, S_t, L_t]', \text{ where}$$

$f(\cdot) \equiv$ production function

Given our assumption concerning the behavior of ϵ , the nonlinear least squares estimates of $\underline{\theta}$ are also the maximum likelihood.

We utilize the modified Gauss-Newton method for finding the values of $\underline{\theta}$ that minimize $S(\underline{\theta})$ (Draper & Smith, 1966, pp. 267-270).³ The estimate of the variance of the errors ϵ_t is

$$s^2 = \frac{1}{T-6} S(\hat{\underline{\theta}}) \text{ where} \quad (8)$$

$\hat{\underline{\theta}} =$ values of $\underline{\theta}$ that minimize equation (6).

²Equation (6) could also be estimated using a Taylor series approximation. See Kmenta (1967). However, the approximation for more than two inputs generates many collinear cross-product terms that make testing for parameter significance difficult. Direct nonlinear maximum likelihood estimation of equation (6) proved the most viable alternative.

³These estimates were also checked by applying the method of estimation known as Marquardt's compromise (Draper & Smith, 1966, p. 272). Little change (except for rounding errors) resulted.

Gallant (1975, p. 74) shows that in large samples, $\hat{\theta}$ has a six-dimensional multivariate normal distribution with mean θ (true value)⁴ and variance-covariance matrix $\sigma^2(F'F)^{-1}$, where F is the $T \times 6$ matrix with elements

$$f_{tj} = \frac{\partial f(\underline{X}_t, \underline{\theta})}{\partial \theta_j} \quad \begin{matrix} t = 1, \dots, T \\ j = 1, \dots, 6. \end{matrix} \quad (9)$$

In hypothesis testing, the matrix $(F'F)^{-1}$ must be approximated by the 6×6 matrix

$$\hat{V} = [F(\hat{\theta})'F(\hat{\theta})]^{-1} \quad (10)$$

A $(1-\alpha)$ percent confidence interval for θ_j may be constructed as

$$\hat{\theta}_j \pm t_{\alpha/2} \sqrt{\hat{S}^2 \hat{d}_{jj}} \quad (11)$$

where $t_{\alpha/2}$ is the $\alpha/2$ critical value of a t -distribution with $T-6$ degrees of freedom and $\hat{d}_{jj}$ is the j th diagonal element of $\hat{V}$. From equation (11) the null hypothesis that $\theta_j = \theta_{j0}$ may be tested at the α percent level of significance by comparing

$$\left| \frac{\hat{\Delta}_{t_j}}{t_j} \right| = \left| \frac{\hat{\theta}_j - \theta_{j0}}{\sqrt{\hat{S}^2 \hat{d}_{jj}}} \right|$$

with $t_{\alpha/2}$ and rejecting the null hypothesis when

$$\left| \frac{\hat{\Delta}_{t_j}}{t_j} \right| > \left| t_{\alpha/2} \right|.$$

⁴On this point, and for a more detailed presentation of the nonlinear least squares estimation technique, see Draper and Smith (1966, pp. 263-304). Also see Gallant (1975) for a concise presentation of hypothesis testing within the nonlinear regression framework.

Results

Equation (6) was estimated for the entire sample (T=227) using the maximum likelihood technique described above.⁵ Regression (1) of Table II indicates an elasticity of substitution (σ) between factors of about 0.8.⁶

⁵One a priori restriction was imposed when estimating equation (6):

$$\sum_{i=1}^3 \delta_i = 1.$$

In this case, each δ_i is interpreted as a "share" of the i th input in the production of scholastic performance. On this point see Ferguson (1969, chapter 5).

⁶The parameters in Table II were calculated from a data matrix where lectures plus seminars (L) is scaled by a factor of 30 and hours studied (S) is scaled by a factor of 100. The differential magnification of these two independent variables facilitated calculation of the standard errors of the regression parameters. To see this, remember that the standard errors are the square roots of the diagonal elements of the matrix $\hat{\sigma}^2 [F(\hat{\theta})' F(\hat{\theta})]^{-1}$ and that

$$f_{t\delta_i} \equiv \frac{\partial f(\underline{x}_t, \underline{\theta})}{\partial \delta_i} \quad (t = 1, \dots, T) \text{ is one of the column vectors}$$

in the $T \times 6$ matrix F . (See equations (8) and (9) of the text for definitions of $\underline{\theta}$, $f(\cdot)$, and F .) Further, notice that for the production function given by equation (6),

$$\partial f(\cdot) / \partial \delta_i = - \frac{MY}{\rho} \left[\cdot \right]^{-\frac{1+\rho}{\rho}} \bar{X}_i^{-\rho} \quad (i = 1, 2, 3).$$

Since ρ is small, the raw values of L and S generate little covariation between $f_{t\delta_1}$ and $f_{t\delta_2}$, and thus lead to a poorly conditioned $F'F$ matrix. When L and S are rescaled, this problem is eliminated.

TABLE II
Nonlinear Least Squares Estimates
CPES Production Functions

	Parameter Estimates ^a						Computed Marginal Products ^b		
	γ	δ_1	δ_2	δ_3	ρ	$\frac{\mu}{\rho}$	$\overline{MP}_L$	$\overline{MP}_S$	$\overline{MP}_C$
(1) <u>The Sample Pooled</u>									
T = 227	6.05	0.10	0.05	0.84	0.24	1.74	0.72	1.04	0.05
SEE = 8.48	(2.61)	(1.26)	(2.62)	(10.90)	(2.25)	(2.09)			
(2) <u>The Sample Stratified</u>									
<u>Males</u>									
T = 128	8.14	0.17	0.04	0.79	0.16	2.32	0.78	0.54	0.04
SEE = 8.51	(1.70)	(1.08)	(1.09)	(4.75)	(0.14)	(0.14)			
(3) <u>Females</u>									
T = 99	4.97	0.00	0.08	0.94	0.19	2.28	0.00	1.24	0.06
SEE = 7.38	(1.57)	(0.00)	(1.09)	(6.19)	(0.09)	(0.09)			

Production Function: $G = \gamma [\delta_1 L^{-\rho} + \delta_2 S^{-\rho} + \delta_3 C^{-\rho}]^{-\frac{1}{\rho}}$

^at-values in parentheses

^b $\overline{MP}_L$, $\overline{MP}_S$, $\overline{MP}_C \equiv$ marginal products evaluated at mean values (unscaled)

For the complete sample, the marginal product of attending one extra lecture or seminar is an increase of about .7 points on the midterm examination. This is 25% less than the marginal product of an hour of study time. Finally, an extra 100 points on the quantitative section of the SAT leads to an exam score higher by approximately 5 points.⁷ From this information we calculate marginal rates of substitution and present them in Table III.

Our results indicate that less able students can indeed compensate for differences in ability with extra study and class attendance. For example, a total of about 7 additional hours of study (about 1½ hours per week for the five week period used here) are necessary to offset a 100 point SAT score disadvantage. If a student were taking four courses with similar learning structures, this result implies that an extra 6 (12) hours per week of study compensate for a 100 (200) point SAT deficiency. Such variations in SAT scores are typical of state universities. The associated study time requirements, while substantial, are not unreasonable. By examining the rest of Table III, one can see the other possible tradeoffs in the production of scholastic performance.

One interesting question that our data permit us to address is whether the possibility of compensating for background deficiency with extra class attendance or study time differs between male and female students. We thus stratify the sample and estimate regression equation (6) separately for males and females. Stratification yields samples of 128 male and 99 female students. Little difference exists between the sexes with respect to μ and ρ , but large differences exist for γ and δ . Females have a smaller γ , implying that on average they score less than males, *ceteris paribus*. The sex differences in parameters in Table III imply that females have a higher marginal product of study and a lower marginal product of class attendance. One possible interpretation of these parameter differences by sex is that the test instrument meters to

⁷The most widely used measures of ability of college students are the verbal and quantitative college board examination, though these scores may capture additional factors. It best served our purpose here to employ the college board quantitative score by itself as an ability proxy, as it is a better independent predictor of exam score than the verbal score alone or the sum of the scores.

TABLE III
 Marginal Rates of Substitution
 Derived from the CPES Production Function
 Estimates of Table II^a

	L	S	C
Total Sample N = 227	-	0.70	13.8
	S	1.43	-
	C	0.07	0.05
			-
Male Sample N = 128	-	1.46	18.1
	S	0.68	-
	C	0.05	0.08
			-
Female Sample N = 99	-	b	b
	S	b	-
	C	b	0.05
			-

^aIn the matrix of marginal rates of technical substitution [S], the element S_{ij} represents the increase in the j^{th} input required to hold output constant when the i^{th} input is decreased by one unit.

^b $MP_L = 0$

some extent memorization ("study") of facts, while class is devoted to concepts. If females are better at memorization than at concept formation (a popular prejudice), and the test weights facts and concepts about equally, then it is reasonable to find that females get relatively more product from study time and males from class time. (We are indebted to an anonymous referee for pointing this out.) Given their higher relative productivity in study, then, efficient allocation of resources within the educational process requires relatively more study time for females. The mean values of S and L in Table I are consistent with this.

CORROBORATION OF EMPIRICAL RESULTS: THE COBB-DOUGLAS FUNCTION

Because of the admitted narrowness of the UNC-CH data, we now examine in some detail the robustness of the empirical results in Table II. We utilize a national data set (Eckland, 1972), which is a detailed follow-up study (performed in 1970) of a group of high school seniors originally surveyed in 1955 by the Educational Testing Service (ETS). The follow-up is a stratified sample of 42 of 516 originally surveyed schools and was selected to provide a proportionate representation of schools across the United States. The Eckland data contain detailed student information on high school courses and performance (including the number of math and science courses taken), an objective measure of ability, and self-ratings of intelligence, diligence, creativity, and intellectual confidence. Student background variables include parental education and income as well as number of siblings and a host of attitudinal questions. Data are also available on freshmen through senior performance (grade-point averages in college and in major field of study).

Despite a broad scope, the Eckland data contain some deficiencies. Certain key variables of particular interest to our research take on only a limited number of values and thus complicate interpretations. For example, ability (APT) is measured as a score on a 20-question test, and quantitative ability is not separated from verbal ability. Only proxies exist for other ability components. Diligence (DILEG) is measured retrospectively on a 4-point scale, and no quantitative information exists for actual time spent studying in and out of class. Given this lack of precise measurement, only qualitative implications with respect to the production of scholastic performance can be drawn. So, rather than employ the computationally cumbersome (expensive) CPES production function, we estimate Cobb-Douglas production

functions via ordinary least squares to check the qualitative aspects of our results in section 4.⁸

Regressions with freshman grade-point average and over-all college grade point average as the dependent variables are presented in the first half of Table IV. These regressions are consistent with our earlier findings that (a) ability and (b) diligence or time intensity of study are key determinants of scholastic performance.⁹

In particular, the coefficients of personal attribute variables are not statistically significant, while those of ability and diligence measures are. Finally, to guard against the possibility that the results in Tables II and IV are similar because of the change in functional form of the regression equation, we estimated the Cobb-Douglas specification with the UNC-CH data. This regression is also presented in Table IV. Ability and diligence factors again dominate background and personal attribute measures as explanatory variables. The results in Table IV illustrate the (qualitative) robustness of the CPES parameter estimates of section 4.

A QUALIFICATION

One qualification is necessary. Learning in one course represents an extremely narrow aspect of the entire learning process within the university. In reality, learning should be viewed as a larger problem in which resources must allocate to the production of grades in a number of competing courses. Viewing learning and time

⁸A logarithmic transformation of the Cobb-Douglas function ($y = \gamma \prod_i x_i^{\delta_i}$) yields an estimating equation

$\ln y = \ln \gamma + \sum \delta_i \ln x_i$. We assume an error term that is independently and normally distributed with mean zero and constant variance. SEX, ATTRAC, AND S*AT are added as control variables.

⁹The Eckland data regressions were also run without an adjustment for high-school grade-point average (GPAHS). Little difference in parameter estimates resulted. The regressions with GPAHS are reported because such a specification is thought to represent a more appropriate scholastic performance production function.

TABLE IV
Scholastic Performance Production Functions
(Cobb-Douglas OLS Specification)

Independent Variables	ECKLAND DATA				UNC-CH DATA			
	Freshman GPA		Overall GPA		Examination Score		Mean Values	Marginal Products
	coef	t-value	coef	t-value	coef	t-value		
<u>Ability:</u>								
CBQ					0.230	4.44	583.53	.033
CBV					0.164	3.32	539.47	.026
APT	0.023	3.70	0.011	2.11				
CREAT	-.066	-2.13	0.016	0.59				
INTEL	0.028	0.55	0.122	2.77				
INTCON	0.085	1.99	0.067	1.82				
GPMS	0.341	8.04	0.257	7.09				
<u>Diligence:</u>								
SEM					0.067	2.21	3.69	1.54
STUDY					0.025	2.13	5.11	.416
LCT					0.034	0.58	8.38	.345
WKST					-.005	-0.44	2.25	-.189
DILEG	0.386	11.53	0.207	7.25				
<u>Background:</u>								
FINC					0.009	0.88	27.52	.028
FED					0.018	0.67	14.61	.105
MED					-.002	-0.06	14.02	-.012
<u>Personal Attributes:</u>								
SEX	-.202	-1.48	-.125	-1.07	-.024	-1.67	0.40	-5.096
ATTRAC	-.068	-1.30	-.028	-0.63				
S*AT	0.070	1.02	-.007	-0.12				
No. Observations	655		655		227			
R ²	0.35		0.30		0.31			

Polachek, Kriesner, and Harwood

TABLE IV (continued)

CBQ	≡ score on college board (SAT) quantitative examination	range: 200-800
CBV	≡ score on college board (SAT) verbal examination	range: 200-800
APT	≡ score on ETS aptitude test	range: 0-20
CREAT	≡ respondent's rating of own creativity	range: 4 (very creative) -1 (not at all creative)
INTEL	≡ respondent's rating of own intelligence	range: 4 (very intelligent) -1 (not at all intelligent)
INTCON	≡ respondent's rating of own intellectual confidence	range: 4 (very) -1 (not at all)
GRAHS	≡ high school grade-point average	range: 1 (F) -5 (A)
SEM	≡ number of seminar sessions attended	
STUDY	≡ number of hours of study for midterm examination	
LCT	≡ number of lectures attended	
WKST	≡ average study hours per week	
DILEG	≡ diligence measured by whether respondent completed class assignments and recommended reading	range: 4 (almost always) -1 (rarely or never)
FINC	≡ family income (\$000's)	
FED	≡ father's level of education	
MED	≡ mother's level of education	
SEX	≡ dummy variable 0 (male) -1 (female)	
ATTRAC	≡ respondent's rating of attractiveness to opposite sex	3 (very) -0 (not at all)
S*AT	≡ interaction of SEX and ATTRAC	

Dependent Variables

Eckland Data: GPA measured on a four point scale

UNC-CH Data: midterm examination score on 100 point scale

Functional Form: n

$$\ln Y = \alpha_0 + \sum_{i=1}^I \alpha_i \ln X_i + \sum_{j=1}^J \beta_j Z_j,$$

where $\ln Y$ is the $\log_e$ of the dependent variable, $\ln X_i$ is the $\log_e$ of the i th input into the production of scholastic performance, and Z_j is the j th conditioning factor.

allocation within the narrow context of one course may lead to biases in our coefficient estimates analogous to simultaneous equations bias. Whereas in our model each input is taken as independent, a more rigorous specification would allow for interrelation between inputs and desired grades across a student's courses. Lack of data in the UNC-CH survey prohibits the construction of such simultaneous models.

SOME POLICY IMPLICATIONS

In this research we view scholastic performance within the context of an individual's production function. We find that innate ability and acquired background are the most important determinants of students' accomplishments. Moreover, time devoted to the learning process proved significant. In particular, class and seminar attendance, as well as study time, yield positive (but differing) marginal products. Estimates of marginal products are potentially useful in allocational decisions within the educational process. For example, suppose that an administrator has X dollars to spend on an activity that improves students' academic performances. Suppose also that by spending this amount on activity A, students each attend one additional class, whereas by spending this amount on activity B, students each study one additional hour. Our hypothetical administrator will make better use of his or her resources by choosing the activity with the higher marginal product.

Our findings also identify some trade-offs within the educational process. For example, we find that approximately $1\frac{1}{2}$ extra hours of study per week compensate for a 100-point college board score deficiency. This result underscores that it is possible for a diligent student with a poor educational background to be academically successful, a fact sometimes neglected by school administrators when formulating admission standards. Finally, our production function estimates illustrate the significance of students' own endeavors in achieving academic success, a factor typically neglected when analyzing the social value of education. Part of the benefits often attributed to public expenditure are direct outputs of student effort.

ACKNOWLEDGMENTS

We wish to thank Mark Appelbaum, Dennis Appleyard, John B. Carroll, William J. Haley, C. Russell Hill, Allen C. Kelley, Thomas Wallsten, and two anonymous referees for their valuable comments. We are especially indebted to Betsy Pierce for her expert help in preparing this manuscript for publication. Versions of this paper have been presented at the ETS-NBER Workshop on the Economics of Education, the Dallas Meetings of the Econometric Society, and the Psychometrics Workshop at the University of North Carolina, Chapel Hill.

REFERENCES

- Bodkin, R. G., & Klein, L. R. Nonlinear estimation of aggregate production functions. Review of Economics and Statistics, 1967, 49, 28-44.
- Bowles, S. Toward an educational production function. In W. L. Hansen, (Ed.), Education, income, and human capital. New York: Columbia University Press for the National Bureau of Economic Research, 1970.
- Brown, B. W., & Saks, D. H. The production and distribution of cognitive skills within schools. Journal of Political Economy, 1975, 83, 571-594.
- Cobb, C. W., & Douglas, P. H. A theory of production. American Economic Review, 1928, 18, 139-165.
- Draper, N. R., & Smith, H. Applied regression analysis. New York: Wiley, 1966.
- Eckland, B. K. Subject index for use with the 1970 survey questionnaire. Working paper no. 2. Chapel Hill: Institute for Research in Social Science, University of North Carolina at Chapel Hill, 1972.
- Eckland, B. K., & MacGillivray, L. School profiles. Working paper no. 1. Chapel Hill: Institute for Research in Social Sciences, University of North Carolina at Chapel Hill, 1972.
- Ferguson, C. C. The neoclassical theory of production and distribution. Cambridge: Cambridge University Press, 1969.

- Gallant, A. R. Nonlinear regression. The American Statistician, 1975, 29, 73-81.
- Gintes, H. Education, technology, and the characteristics of worker productivity. American Economic Review Proceedings, 1971, 61, 266-279.
- Hanushek, E. Teacher characteristics and gains in student achievement: estimation using micro data. American Economic Review, 1971, 61, 280-288.
- Kmenta, J. On the estimation of the CES production function. International Economic Review, 1967, 8, 180-189.
- Kumar, T. K., & Capinski, J. H. Nonlinear estimation of the CES production function: sampling distributions and tests in small samples. Southern Economic Journal, 1974, 41, 258-266.
- Lavin, D. E. The prediction of academic performance. New York: Russell Sage Foundation, 1965.
- Lydall, H. The structure of earnings. Oxford: Clarendon, 1968.
- Meyer, R. A., & Kadiyala, K. R. Linear and nonlinear estimation of production function. Southern Economic Journal, 1974, 40, 463-472.
- Mincer, J. Schooling, experience, and earnings. New York: Columbia University Press for the National Bureau of Economic Research, 1974.
- Wise, D. A. Academic achievement and job performance. American Economic Review, 1975, 65, 350-366.
- Uzawa, H. Production functions with constant elasticities of substitution. Review of Economic Studies, 1962, 24, 291-299.

AUTHORS

POLACHEK, SOLOMON W. Address: Department of Economics,
The University of North Carolina at Chapel Hill,
Gardner Hall 017-A, Chapel Hill, North Carolina, 27514.
Title: Associate Professor. Degrees: A.B., The George
Washington University; Ph.D., Columbia University.
Specialization: labor economics.

KNIESNER, THOMAS J. Address: Department of Economics,
The University of North Carolina at Chapel Hill,
Gardner Hall 017-A, Chapel Hill, North Carolina, 27514.
Title: Assistant Professor. Degrees: B.A., M.A.,
Ph.D., The Ohio State University. Specialization:
labor economics.

HARWOOD, HENRICK J. Address: Department of Economics,
The University of North Carolina at Chapel Hill,
Gardner Hall 017-A, Chapel Hill, North Carolina, 27514.
Title: Graduate Student. Degrees: B.A., Stetson
University. Specialization: economics of education.